

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/31, C07K 14/315, C12N 1/21, C07K 16/12, A61K 39/09, 39/395, G01N 33/569, C12Q 1/68	A1	(11) International Publication Number: WO 96/40928 (43) International Publication Date: 19 December 1996 (19.12.96)
(21) International Application Number: PCT/CA96/00322 (22) International Filing Date: 17 May 1996 (17.05.96) (30) Priority Data: 08/472,534 7 June 1995 (07.06.95) US 60/001,805 4 August 1995 (04.08.95) US (71) Applicant (for all designated States except US): IAF BIOVAC INC. [CA/CA]; 525 Des Prairies Boulevard, Laval, Quebec H7N 4Z2 (CA). (72) Inventors; and (75) Inventors/Applicants (for US only): HAMEL, Josée [CA/CA]; (CA). BRODEUR, Bernard [CA/CA]; 2401 Maritain Street, Sillery, Quebec G1T 1N6 (CA). MARTIN, Denis [CA/CA]; 4728-G Gaboury Street, St-Augustin-de-DesMaures, Quebec G3A 2X1 (CA). RIOUX, Clément [CA/CA]; 1012 Jean Charles Cantin, Ville de Cap-Rouge, Quebec G1Y 2X1 (CA). (74) Agents: DUBUC, Jean, Y. et al.; The Stock Exchange Tower, Suite 3400, 800 Square Victoria, P.O. Box 242, Montreal, Quebec H4Z 1E9 (CA).		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published With international search report.
(54) Title: STREPTOCOCCAL HEAT SHOCK PROTEINS MEMBERS OF THE HSP70 FAMILY (57) Abstract Novel heat shock proteins (HSPs) of <i>Streptococcus pneumoniae</i> , <i>Streptococcus pyogenes</i> and <i>Streptococcus agalactiae</i> having apparent molecular masses of 70-72 kDa, immunologically related polypeptides, the nucleotide and derived amino acid sequences of HSP72 of <i>S. pneumoniae</i> (SEQ ID NO:4; SEQ ID NO:5), the nucleotide and derived amino acid sequences of HSP70 of <i>S. pyogenes</i> (SEQ ID NO:19; SEQ ID NO:20), the nucleotide and derived amino acid sequences of HSP 70 of <i>S. agalactiae</i> (SEQ ID NO:21; SEQ ID NO:22), antibodies that bind to the HSPs, and recombinant DNA methods for the production of the HSPs and immunologically related polypeptides are described. The polypeptides, DNA sequences and antibodies of this invention provide new means for the diagnosis, prevention and/or treatment of Streptococcal disease.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

**STREPTOCOCCAL HEAT SHOCK PROTEINS
MEMBERS OF THE HSP70 FAMILY**

5 **TECHNICAL FIELD OF THE INVENTION**

This invention relates to novel heat shock proteins of *Streptococcus pneumoniae*, *Streptococcus pyogenes* and *Streptococcus agalactiae* and immunologically related polypeptides, which provide the basis for new immunotherapeutic, prophylactic and diagnostic agents useful in the treatment, prevention and diagnosis of disease. More particularly, this invention relates to heat shock proteins of *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*, members of the HSP70 family which have an apparent molecular mass of 70-72 kilodaltons, to the corresponding nucleotide and derived amino acid sequences, to recombinant DNA methods for the production of HSP70/HSP72 and immunologically related polypeptides, to antibodies that bind to these HSP's, and to methods and compositions for the diagnosis, prevention and treatment of diseases caused by *S. pneumoniae* and related bacteria, such as *Streptococcus pyogenes* and *Streptococcus agalactiae*

25

BACKGROUND OF THE INVENTION

S. pneumoniae is an important agent of disease in humans, especially among infants, the elderly and immunocompromised persons. It is a bacterium frequently isolated from patients with invasive diseases such as bacteraemia/septicaemia, pneumonia, and meningitis with high morbidity and mortality throughout the world. Although the advent of antimicrobial drugs has reduced the overall mortality from pneumococcal diseases, the presence of resistant pneumococcal organisms has become a major problem in the world today. Effective pneumococcal vaccines could have a major impact on the morbidity and mortality associated with *S. pneumoniae* disease. Such

vaccines would also potentially be useful to prevent otitis media in infants and young children.

It is clear that a number of pneumococcal factors are potentially important in the pathogenesis of disease [G.J. Boulnois, J. Gen. Microbiol., 138, pp. 249-259 (1992); C.J. Lee et al., Crit. Rev. Microbiol., 18, pp. 89-114 (1991)]. The capsule of the pneumococcus, despite its lack of toxicity, is considered to be the *sine qua non* of pneumococcal virulence. More than 80 pneumococcal capsular serotypes are identified on the basis of antigenic differences. Antibodies are the mechanism of protection and the importance of anticapsular antibodies in host defenses against *S. pneumoniae* is well established [R. Austrian, Am. J. Med., 67, pp. 547-549 (1979)]. Nevertheless, the currently available pneumococcal vaccine, comprising 23 capsular polysaccharides that most frequently caused disease, has significant shortcomings such as the poor immunogenicity of capsular polysaccharides, the diversity of the serotypes and the differences in the distribution of serotypes over time, geographic areas and age groups. In particular, the failure of existing vaccines to protect young children against most serotypes has spurred evaluation of other *S. pneumoniae* components. Increasing evidence indicates that certain pneumococcal proteins may play an active role both in terms of protection and pathogenicity [J.C. Paton, Ann. Rev. Microbiol., 47, pp. 89-115 (1993)]. So far, however, only a few *S. pneumoniae* proteins have been studied. This might result from the lack of protein-specific antibodies which renders difficult the study of the role of protein antigens in protection and pathogenicity. It is believed that the pneumococcal protein antigens are not very immunogenic and that most antibody responses are to the phosphocholine and the capsular polysaccharides [L.S. McDaniel et al., J. Exp. Med., 160, pp. 386-397 (1984); R.M. Krause, Adv. Immunol., 12, pp. 1-56 (1970); D.G. Braun et al., J. Exp.

Med., 129, pp. 809-830 (1969)]. In a study using X-linked immunodeficient mice, which respond poorly to carbohydrate antigens and to phosphocholine, but make relatively normal responses to protein antigens, the frequency for obtaining
5 monoclonal antibodies reactive with pneumococcal protein antigens was less than 10%, thus suggesting that *S. pneumoniae* proteins are poor immunogens [McDaniel et al., supra].

Streptococcus agalactiae, also called Group B
10 *Streptococcus* (GBS), is the most common cause of sepsis (blood infection) and meningitis in newborns. GBS is also a frequent cause of newborn pneumonia. Approximately 8,000 babies in the United States get GBS disease each year; 5%-15% of these babies die. Babies that survive,
15 particularly those who have meningitis, may have long-term problems, such as hearing or vision loss or learning disabilities. In pregnant women, GBS can cause urinary tract infections, womb infections (amnionitis, endometritis), and stillbirth. Among women who are not
20 pregnant and men, the most common diseases caused by GBS are blood infections, skin or soft tissue infections, and pneumonia. Approximately 20% of men and nonpregnant women with GBS disease die of the disease. GBS infections in both newborns and adults are usually treated with
25 antibiotics (e.g., penicillin or ampicillin) given intravenously. Most GBS disease in newborns can be prevented by giving certain pregnant women antibiotics intravenously during labor. Vaccines to prevent GBS disease are being developed. In the future, it is
30 expected that women who will be vaccinated will make antibodies that cross the placenta and protect the baby during birth and early infancy.

Since the 1980s, *Streptococcus pyogenes*, also called Group A *Streptococcus* (GAS) is reemerging as a
35 cause of severe diseases which would be due to an increase

in virulence of the organism. GAS causes pharyngitis, commonly called "strep throat", and skin infections (impetigo, erysipelas/cellulitis). "Strep throat" and impetigo can lead to glomerulonephritis (kidney damage).

5 Approximately 3% of "strep throat" infections result into rheumatic fever (migrating arthritis) whose complications include chorea (neurological symptoms) and, in 50% of the cases, rheumatic heart disease (heart valve damage) with endocarditis as a possible long term consequence. It is

10 important to treat impetigo and "strep throat" with antibiotics to prevent the development of complications. Infection with toxin-producing strains can result in scarlet fever (diffuse rash and fever) or in the extremely severe streptococcal toxic shock syndromes (TSS; GAS have

15 been termed 'flesh eating bacteria') which are characterized by the rapid development of shock and multiple organ system failure. TSS have a 30 to 70% fatality rate in spite of aggressive treatment involving the removing of the focus of bacterial infection and

20 antibiotic therapy. The incidence of TSS is 10 to 20 cases per 100,000. No vaccine against GAS is presently available.

Heat shock or stress proteins ("HSPs") are among the most highly conserved and abundant proteins found in

25 nature [F.C. Neidhardt et al., Ann. Rev. Genet., 18, pp. 295-329 (1984); S. Lindquist, Ann. Rev. Biochem., 55, pp. 1151-1191 (1986)]. They are produced by all cells in response to various physiological and nonphysiological stimuli. The heat shock response, in which a sudden

30 increase in temperature induces the synthesis of HSPs, is the best studied of the stress responses. Other environmental conditions such as low pH, iron deficiency and hydrogen peroxyde can also induce HSPs. The HSPs have been defined by their size, and members of hsp90, hsp70,

35 and hsp60 families are among the major HSPs found in all prokaryotes and eukaryotes. These proteins fulfill a

variety of chaperon functions by aiding protein folding and assembly and assisting translocation across membranes [C. Georgopoulos and W.J. Welch, Ann. Rev. Cell. Biol., 9, pp. 601-634 (1993); D. Ang et al., J. Biol. Chem., 266, pp. 24233-24236 (1991)]. As molecular chaperons and possibly via other mechanisms, HSPs are likely involved in protecting cells from the deleterious effects of stress. The fact that several virulence factors are regulated by environmental conditions suggests a role for HSPs in microbial pathogenicity [J.J. Mekalanos, J. Bacteriol., 174, pp. 1-7 (1992); P.J. Murray and R.A. Young, J. Bacteriol., 174, pp. 4193-4196 (1992)]. In that respect, recent studies on *Salmonella* species suggest that the stress response might be critically linked to the ability of intracellular pathogens to initiate and sustain an infection [N.A. Buchmeir and F. Heffron, Science, 248, pp. 730-732 (1990); K.Z. Abshire and F.C. Neidhardt, J. Bacteriol., 175, pp. 3734-3743 (1993); B.B. Finlay et al., Science, 243, pp. 940-943 (1989)]. Others have demonstrated that lysteriolysin, an essential virulence factor in *L. monocytogenes*, is induced under heat shock conditions [Z. Sokolovic and W. Goebel, Infect. Immun., 57, pp. 295-298 (1989)].

Evidence is now accumulating that HSPs are major antigens of many pathogens. Members of the hsp60 family, also called GroEL-related proteins for their similarity to the *E. coli* GroEL protein, are major antigens of a variety of bacterial pathogens including *Mycobacterium leprae* and *Mycobacterium tuberculosis* [D. Young et al., Proc. Natl. Acad. Sci. USA, 85, pp. 4267-4270 (1988)], *Legionella pneumophila* [B.B. Plikaytis et al., J. Clin. Microbiol., 25, pp. 2080-2084 (1987)], *Borrelia burgdorferi* [B.J. Luft et al., J. Immunol., 146, pp. 2776-2782 (1991)], and *Chlamydia trachomatis* [E.A. Wagar et al., J. Infect. Dis., 162, pp. 922-927 (1990)]. This antigen is a homologue of the ubiquitous "common antigen", and is believed to be present in every bacterium [J.E. Thole et al., Microb.

Pathogen., 4, pp. 71-83 (1988). Antibodies to the members of the hsp70 family, or DnaK-related proteins, have also been described for several bacterial and parasitic infections [Young et al., supra; Luft et al., supra; D.M. Engman et al., J. Immunol., 144, pp. 3987-3991 (1990); N.M. Rothstein et al., Molec. Biochem. Parasitol., 33, pp. 229-235 (1989); V. Nussenzweig and R.S. Nussenzweig, Adv. Immunol., 45, pp. 283-334 (1989)]. HSPs can elicit strong B- and T- cell responses and it was shown that 20% of the CD4⁺ T-lymphocytes from mice inoculated with *M. tuberculosis* were reactive to the hsp60 protein alone [S.H.E. Kaufman et al., Eur. J. Immunol., 17, pp. 351-357 (1987)]. Similarly, 7 out of a collection of 24 monoclonal antibodies to *M. leprae* proteins recognized determinants on hsp60 [H.D. Engers et al., Infect. Immun., 48, pp. 603-605 (1985)]. It seems that the immune response to stress proteins might play an important role in protection against infection. Consistent with that is the demonstration that antibodies and T cells reactive with microbial HSPs can exhibit neutralizing and protective activities [A. Noll et al., Infect. Immun., 62, pp. 2784-2791 (1994); and S.L. Danilition et al., Infect. Immun., 58, pp. 189-196 (1990)]. The immunological properties of stress proteins make them attractive as vaccine components and several HSPs are presently being considered for preventing microbial infection and treating cancer. So far, however, studies have focused on intracellular pathogens such as *Mycobacteria*, *Salmonella*, *Chlamydia* and several parasites. Information concerning the heat shock protein antigens in extracellular gram-positive bacteria is far less documented. In *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*, neither the heat shock proteins nor their gene structures have been identified.

DISCLOSURE OF THE INVENTION

The present invention addresses the problems referred to above by providing novel heat shock proteins

from *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*, and immunologically related polypeptides. Also provided are DNA sequences that code for the foregoing polypeptides, vectors containing the polypeptides, unicellular hosts
5 transformed with those vectors, and a process for making substantially pure, recombinant polypeptides. Also provided are antibodies specific to the foregoing polypeptides. The polypeptides, DNA sequences and antibodies of this invention provide the basis for novel
10 methods and pharmaceutical compositions for the detection, prevention and treatment of disease. Particularly, this invention provides a novel vaccine based on fragments of these polypeptides that are specific to streptococcal strains.

15 The novel heat shock protein is the approximately 72 kDa heat shock protein of *Streptococcus pneumoniae* ("HSP72") (SEQ ID NO:5), the approximately 70 kDa heat shock protein of *Streptococcus pyogenes* ("HSP70") (SEQ ID NO:20) and the approximately 70 kDa heat shock
20 protein of *Streptococcus agalactiae* ("HSP70") (SEQ ID NO:22), including analogues, homologues, and derivatives thereof, and fragments of the foregoing polypeptides containing at least one immunogenic epitope. Preferred fragments of HSP70/72 include the C-terminal portion of
25 the HSP70/72 polypeptides. More particularly, it includes the C-terminal 169-residue fragment ("C-169") (residues 439-607, SEQ ID NO:5), the C-terminal 151-residue fragment ("C-151") (residues 457-607, SEQ ID No:5), and smaller fragments consisting of peptide epitopes within the C-169
30 region. Particularly preferred fragments within the C-169 region of HSP72 include the peptide sequences GFDAERDAAQAALDD (residues 527-541 of SEQ ID NO:5) and AEGAQATGNAGDDVV (residues 586-600 of SEQ ID NO:5), which are exclusive to HSP72 of *Streptococcus pneumoniae*. Even
35 more preferred are fragments that elicit an immune reaction against *S. pneumoniae*, *S. pyogenes* and *S.*

agalactiae but do not provoke auto-immune reaction in a human host. Such fragments may be selected from the following peptides: CS870, CS873, CS874, CS875, CS876, CS877, CS878, CS879, CS880, CS882, MAP1, MAP2, MAP3 and
5 MAP4 (see TABLE 5, supra).

Preferred antibodies of this invention are the F1-Pn3.1, F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4 monoclonal antibodies ("MAbs"), which are specific to HSP72.

More preferred antibodies are the F2-Pn3.2 and
10 F2-Pn3.4 monoclonal antibodies that are specific to both HSP 70 and HSP72. Even more preferred are the F1-Pn3.1 antibodies that are specific for *Streptococcus pneumoniae*.

The preferred polypeptides and antibodies of this invention provide the basis for novel methods and
15 pharmaceutical compositions for the detection, prevention and treatment of pneumococcal diseases.

BRIEF DESCRIPTION OF THE DRAWINGS

20 FIG. 1 depicts a fluorogram, which shows the effect of heat shock on *S. pneumoniae* protein synthesis. The cell extracts in panel A are *S. pneumoniae* type 6 strain 64. The cell extracts in panel B are *S. pneumoniae* type 4 strain 53. The cell extracts in the odd numbered
25 lanes were incubated at 37°C. The cell extracts in the even numbered lanes were incubated at 45°C for 5 minutes. The cell extracts were then labeled with [³⁵S]methionine for 10 minutes (lanes 1, 2 and 7, 8), 30 minutes (lanes 3, 4 and 9, 10), or 60 minutes (lanes 5, 6). Molecular mass
30 markers in kilodaltons are shown to the left. The positions of HSP80, HSP72 and HSP62 are shown by arrows at the right-hand side of each panel.

FIG. 2 is a graphical depiction of a comparison of the electrophoretic profiles of [³⁵S]methionine-labeled
35 proteins in *S. pneumoniae* in the presence (----) or absence (____) of exposure to heat shock. Densitometric tracings were determined by measuring the relative optical

density (Y axis) vs. the mobility of labeled protein bands (X axis). The densitometric scans of the SDS PAGE of FIG. 1, lanes 1 and 2, is shown.

FIG. 3 depicts a fluorogram, which shows the
5 *S. pneumoniae* protein antigens immunoprecipitated by sera from mice immunized with detergent-soluble *S. pneumoniae* protein extract. [³⁵S]methionine-labeled proteins from *S. pneumoniae* grown at 37°C and incubated at 37°C (lanes 3, 5, 7 and 9) or heat-shocked at 45°C (lanes 4, 6, 8 and 10)
10 were immunoprecipitated with sera from mouse 1 (lanes 3 to 6) or mouse 2 (lanes 7 to 10) and then analyzed by SDS-PAGE and fluorography. The sera were tested after the first (lanes 3,4 and 7,8) and after the second (lanes 5,6 and 9,10) immunization. Cell lysates from [³⁵S]methionine-
15 labeled non heat-shocked and heat-shocked *S. pneumoniae* are shown in lanes 1 and 2, respectively. The position of HSPs is indicated by the arrows at the left of the fluorogram.

FIG. 4 depicts a fluorogram, which shows the
20 *S. pneumoniae* protein antigens immunoprecipitated by sera from mice immunized with heat-killed *S. pneumoniae* bacteria. [³⁵S]methionine-labeled proteins from *S. pneumoniae* grown at 37°C and incubated at 37°C (lanes 3, 5 and 7) or heat-shocked at 45°C (lanes 4, 6 and 8) were
25 immunoprecipitated with sera from mouse 1 (lanes 3,4), mouse 2 (lanes 5,6) or mouse 3 (lanes 7, 8) and then analyzed by SDS-PAGE and fluorography. Sera were tested after the second immunization only. Cell lysates from [³⁵S]methionine-labeled non heat- and heat-shocked
30 *S. pneumoniae* are shown in lanes 1 and 2, respectively. The position of HSPs is indicated by the arrows at the left of the fluorogram.

FIG. 5 depicts a photograph, which shows the
S. pneumoniae antigens detected by Western blot analysis.
35 Whole cell extracts were probed with sera from 15 mice (lanes 1-15) immunized with heat-killed *S. pneumoniae* bacteria. Lane 16 shows the HSP72 protein detected by MAb

F1-Pn3.1. In panel A, the sera were tested after the second immunization. In panel B, the reactivity of 4 out of 15 sera tested after the first immunization is shown. The positions of 53.5 kDa- and 47 kDa-protein bands are indicated by the bars at the left. The position of HSP72 is shown by the arrows at the right of each panel.

FIG. 6 depicts a fluorogram showing the specificity of MAb F1-Pn3.1 for HSP72. [³⁵S]methionine-labeled proteins of *S. pneumoniae* in the absence (lanes 1, 3 and 5) or presence (lanes 2, 4 and 6) of exposure to heat shock were immunoprecipitated with IgG2a-control MAb (lane 3,4) or F1-Pn3.1 (lane 5,6) and then analyzed by SDS-PAGE and fluorography. Cell lysates from [³⁵S]methionine-labeled non heat-shocked and heat-shocked *S. pneumoniae* are shown in lanes 1 and 2, respectively. The position of HSPs (all three) is shown by the arrows at the left of the fluorogram.

FIG. 7, panel A, depicts an immunoblot, which shows the reaction of heat-shocked and non heat-shocked [³⁵S]methionine-labelled *S. pneumoniae* cell extracts with MAb F1-Pn3.1. Lane 1 contains heat-shocked cell lysates (45°C). Lane 2 contains non heat-shocked cell lysates (37°C). Panel B depicts a fluorogram of the immunoblot shown in panel A.

FIG. 8 depicts a Western Blot, which shows subcellular localization of *S. pneumoniae* HSP72. Sample containing 15 µg protein of membrane fraction (lane 1) and cytoplasmic fraction (lane 2) of *S. pneumoniae* were electrophoresced on SDS-PAGE transferred to nitrocellulose and probed with MAb F1-Pn3.1.

FIG. 9 is a photograph of an immunoblot showing the reactivity of recombinant fusion proteins containing the C-169 region of *S. pneumoniae* HSP72 with MAb F1-Pn3.1. Lane 1 contains whole cell extracts from *S. pneumoniae* strain 64 probed with HSP72-specific MAb F1-Pn3.1. Lanes 2 and 3 contain phage lysates from *E. coli* infected with λJBD17 cultured in the presence (+) or absence (-) of

IPTG and probed with HSP72-specific MAb F1-Pn3.1. Lanes 4 and 5 contain phage lysates from *E. coli* infected with λ JBD7 cultured in the presence (+) or absence (-) of IPTG and probed with HSP72-specific MAb F1-Pn3.1. Molecular mass markers are shown to the left. The positions of the 74kDa- and 160 kDa-reactive proteins are shown on the left and on the right, respectively.

FIG. 10 is a schematic representation of the restriction map of the HSP72(DnaK) and *Fuc* loci and inserts of recombinant clones. The relationships between DNA fragments are shown with respect to each other. FIGS. 10A and 10C illustrate the restriction map of the HSP72(DnaK) and *Fuc* loci, respectively. FIG 10B illustrates the inserts of the various phages and plasmids described in Example 3. H(HindIII); E(EcoRI); V(EcoRV); P(PstI); and X(XhoI) indicate positions of restriction endonuclease sites. DNA fragments on the HSP72/DnaK locus (■); the *Fuc* locus (///); and fragments used as probes in the Southern blot analyses (▤) are indicated.

FIG. 11 depicts the SDS-PAGE and Western blot analyses of the recombinant 74 kDa protein. Whole cell extracts from *E. coli* transformed with plasmids pJBD179 (lane 1), pJBDF51 (lanes 2 and 3) and pJBDF62 (lane 4 and 5) and cultured in presence (+) or absence (-) of IPTG were subjected to 10% polyacrylamide gel electrophoresis. The proteins were then visualized by Coomassie Blue staining (A) or Western blotting (B) using HSP-specific MAb F1-Pn3.1. Molecular mass markers in kilodaltons are shown to the left. The arrow at the left-hand side of each panel marks the 74 kDa protein marker.

FIG. 12 depicts the detection of native and recombinant HSP72 antigens by Western blot analysis. Whole cell lysates from *E. coli* transformed with plasmids pJBDF51 (lanes 1 and 3) and pJBD291 (lane 2) and cell lysates from *S. pneumoniae* strain 64 (lane 4) were subjected to 10% polyacrylamide gel electrophoresis and

were electrotransferred to nitrocellulose. The immunoblot was probed with HSP72-specific MAb F1-Pn3.1.

FIGS. 13A-13D depict a comparison of the predicted amino acid sequence of the *S. pneumoniae* HSP72 open reading frame (HSP72 SPNEU) with those previously reported for the following HSP70/DnaK proteins: ECOLI, *Escherichia coli*; BORBU, *Borrelia burgdorferi*; BRUOV, *Brucella ovis*; CHLPN, *Chlamydia pneumoniae*; BACME, *Bacillus megaterium*; BACSU, *Bacillus subtilis*; STAAU, *Staphylococcus aureus*; LACLA, *Lactococcus lactis*; and MYCTU, *Mycobacterium tuberculosis*. Only mismatched amino acids are indicated. Identical and conserved amino acids are boxed and shadowed, respectively.

FIG. 14 depicts a photograph of an SDS-PAGE, which shows the recombinant *S. pneumoniae* HSP72 purified by affinity chromatography. Supernatant fractions from *E. coli* (pJBDk51) lysates (lane 2) and 20 µg of immunoaffinity-purified HSP72_{rec} (lane 3) were subjected to 10% polyacrylamide gel electrophoresis. The proteins were then visualized by Coomassie Blue staining. Lane 1 shows the migration of molecular mass markers (106 kDa, 80 kDa, 49.5 kDa, 32.5 kDa, 27.5 kDa and 18.5 kDa).

FIG. 15 depicts a photograph of SDS-PAGE, which shows the recombinant *S. pneumoniae* C-169 fragment purified by solubilization of inclusion bodies. Various amounts of purified C-169 protein (lane 1, 5 µg; lane 2, 2.5 µg; and lane 3, 1 µg) and whole cell lysates from *E. coli* transformed with plasmids pDELTA1 (lane 4) and pJBDA1 (lane 5) were subjected to 10% polyacrylamide gel electrophoresis. The proteins were then visualized by Coomassie Blue staining.

FIG. 16 is a graphical depiction of the survival curve of Balb/c mice protected from *S. pneumoniae* infection by immunization with HSP72_{rec}. Data are presented as the per cent (%) survival over a period of 14 days for a total of 10 mice per experimental group.

FIG. 17 is a graphical depiction of the survival curve of Balb/c mice protected from *S. pneumoniae* infection by immunization with C-169_{rec}. Data are presented as the per cent (%) survival over a period of 14 days for a total of 10 mice per experimental group.

FIG. 18 is a map of plasmid pURV3 containing C-151_{rec}, the coding region for the 151 amino acids at the carboxyl end of the HSP72 of *S. pneumoniae*; Amp^R, ampicillin-resistance coding region; ColE1 ori, origin of replication; cI857, bacteriophage λ cI857 temperature-sensitive repressor gene; λ PL, bacteriophage λ transcription promoter; T1, T1 transcription terminator. The direction of transcription is indicated by the arrows. *Bgl*III and *Bam*HI are the restriction sites used to insert the coding region for the C-151_{rec} of the HSP72 of *S. pneumoniae*.

FIG. 19 illustrates the distribution of anti-*S. pneumoniae* titers in sera from Balb/c mice immunized with HSP72_{rec}. Sera were collected after the first, second and third injection with 1 μ g (O) or 5 μ g (●) of HSP72_{rec} and evaluated individually for anti-*S. pneumoniae* antibody by ELISA. Titers were defined as the highest dilution at which the A410 values were 0.1 above the background values. Plain lines indicate the median reciprocal of antibody titers for each group of mice while the dashed line indicates the median value for preimmune sera.

FIG. 20 illustrates the distribution of anti-*S. pneumoniae* titers in sera from Balb/c mice immunized with C-169_{rec}. Sera were collected after the first, second and third injection with 1 μ g (O) or 5 μ g (●) of C-169_{rec} and evaluated individually for anti-*S. pneumoniae* antibody by ELISA. Titers were defined as the highest dilution at which the A410 values were 0.1 above the background values. Plain lines indicate the median reciprocal of antibody titers for each group of mice while the dashed line indicates the median value for preimmune sera.

FIG. 21 illustrates the distribution of anti-*S. pneumoniae* titers in sera from Balb/c mice immunized with C-151_{rec}. Sera were collected after the first, second and third injection with 0.5 µg of C-151_{rec} and evaluated individually for anti-*S. pneumoniae* antibody by ELISA. Titters were defined as the highest dilution at which the A410 values were 0.1 above the background values. Plain lines indicate the median reciprocal of antibody titers for each group of mice while the dashed line indicates the median value for preimmune sera.

FIG. 22 illustrates the antibody response of cynomolgus monkeys immunized with recombinant HSP72 antigens. Groups of two monkeys were immunized with either HSP72_{rec} or C-169_{rec} protein at day 1, day 22 and day 77. Sera were collected regularly during the course of the immunization and evaluated individually for pneumococcal HSP72 specific antibody by Western blot analysis. Titters were defined as the highest dilution at which the HSP72 band was visualized.

FIG. 23 illustrates the binding of hyperimmune sera to peptides in a solid-phase ELISA. Rabbit, mouse and monkey sera from animals immunized with either HSP72_{rec} or C-169_{rec} protein were tested for their reactivity to peptides. Optical density values were obtained with sera tested at a dilution of 1:100 except for the values corresponding to the reactivity of rabbit sera to peptide MAP2 and murine sera to peptides MAP2 and MAP4 which were obtained with sera diluted 1:1000.

FIG. 24 depicts the consensus sequence established from the DNA sequences of the *hsp70/dnak* open reading frames of *Streptococcus pneumoniae* (spn-orf), *Streptococcus pyogenes* (sga-orf) and *Streptococcus agalactiae* (sgb-orf) and indicates the substitutions and insertions of nucleotides specific to each species.

FIG. 25 depicts the consensus sequence established from the protein sequences of the Hsp70 of *Streptococcus pneumoniae* (spn-prot), *Streptococcus pyogenes* (sga-prot)

and *Streptococcus agalactiae* (sgb-prot) and indicates the substitutions and insertions of amino acids specific to each species.

FIG. 26 depicts a fluorogram, which shows the effect of heat shock on *S. agalactiae* protein synthesis and the *S. agalactiae* protein antigen immunoprecipitated by MAb F2-Pn3.4. Cell lysates from [³⁵S]methionine-labeled proteins from *S. agalactiae* grown at 37°C and incubated at 37°C (odd numbered lanes) or heat-shocked at 43°C (even numbered lanes) were analysed by SDS-PAGE and fluorography. Lanes 3 and 4 show the immunoprecipitates obtained using MAb F2-Pn3.4.

DETAILED DESCRIPTION OF THE INVENTION

According to one aspect of the invention, we provide novel heat shock proteins of *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*, and analogues, homologues, derivatives and fragments thereof, containing at least one immunogenic epitope. As used herein, a "heat shock protein" is a naturally occurring protein that exhibits preferential transcription during heat stress conditions. The heat shock protein according to the invention may be of natural origin, or may be obtained through the application of recombinant DNA techniques, or conventional chemical synthesis techniques.

As used herein, "immunogenic" means having the ability to elicit an immune response. The novel heat shock proteins of this invention are characterized by their ability to elicit a protective immune response against Streptococcal infections, more particularly against lethal *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*.

The invention particularly provides a *Streptococcus pneumoniae* heat shock protein of approximately 72 kDa ("HSP72"), having the deduced amino acid sequence of SEQ ID NO:5, and analogues, homologues, derivatives and

fragments thereof, containing at least one immunogenic epitope.

As used herein, "analogues" of HSP72 are those *S. pneumoniae* proteins wherein one or more amino acid residues in the HSP72 amino acid sequence (SEQ ID NO:5) is replaced by another amino acid residue, providing that the overall functionality and immunogenic properties of the analogue protein are preserved. Such analogues may be naturally occurring, or may be produced synthetically or by recombinant DNA technology, for example, by mutagenesis of the HSP72 sequence. Analogues of HSP72 will possess at least one antigen capable of eliciting antibodies that react with HSP72, e.g. *Streptococcus pyogenes* and *Streptococcus agalactiae*.

As used herein, "homologues" of HSP72 are proteins from Streptococcal species other than *pneumoniae*, *pyogenes* or *agalactiae*, or genera other than *Streptococcus* wherein one or more amino acid residues in the HSP72 amino acid sequence (SEQ ID NO:5) is replaced by another amino acid residue, providing that the overall functionality and immunogenic properties of the homologue protein are preserved. Such homologues may be naturally occurring, or may be produced synthetically or by recombinant DNA technology. Homologues of HSP72 will possess at least one antigen capable of eliciting antibodies that react with HSP72, e.g. *Enterococcus faecalis*.

As used herein, a "derivative" is a polypeptide in which one or more physical, chemical, or biological properties has been altered. Such alterations include, but are not limited to: amino acid substitutions, modifications, additions or deletions; alterations in the pattern of lipidation, glycosylation or phosphorylation; reactions of free amino, carboxyl, or hydroxyl side groups of the amino acid residues present in the polypeptide with other organic and non-organic molecules; and other alterations, any of which may result in changes in primary, secondary or tertiary structure.

The "fragments" of this invention will have at least one immunogenic epitope. An "immunogenic epitope" is an epitope that is instrumental in eliciting an immune response. The preferred fragments of this invention will
5 elicit an immune response sufficient to prevent or lessen the severity of infection, e.g., *S. pneumoniae* infection. Preferred fragments of HSP72 include the C-terminal region of the polypeptides. More preferred fragment include the C-terminal 169-residue fragment ("C-169") (SEQ ID NO:5,
10 residues 439-607), the C-terminal 151-residue ("C-151") (SEQ ID No:5, residues 457-607) and smaller fragments consisting of peptide epitopes within the C-169 region. Particularly preferred fragments within the C-169 region of HSP72 include the peptide sequences GFDAERDAAQAALDD
15 (residues 527-541 of SEQ ID NO:5) and AEGAQATGNAGDDVV (residues 586-600 of SEQ ID NO:5), which are exclusive to HSP72 of *Streptococcus pneumoniae*, or corresponding degenerate fragments from *S. pyogenes* or *S. agalactiae* (see FIG. 25). Even more preferred are fragments that
20 elicit a specific immune reaction against *Streptococcal* strains. Such fragments may be selected from the following peptides: CS870, CS873, CS874, CS875, CS876, CS877, CS878, CS879, CS880, CS882, MAP1, MAP2, MAP3 and MAP4 (see TABLE 5, supra), or homologues thereof.

25 In a further aspect of the invention, we provide polypeptides that are immunologically related to HSP70/72. As used herein, "immunologically related" polypeptides are characterized by one or more of the following properties:

(a) they are immunologically reactive with
30 antibodies generated by infection of a mammalian host with *Streptococcus pneumoniae* cells, which antibodies are immunologically reactive with HSP72 (SEQ ID NO:5) and HSP70 (SEQ ID NO:20 and SEQ ID NO:22);

(b) they are capable of eliciting antibodies that
35 are immunologically reactive with HSP72 (SEQ ID NO:5) and HSP70 (SEQ ID NO:20 and SEQ ID NO:22);

(c) they are immunologically reactive with antibodies elicited by immunization of a mammal with HSP72 (SEQ ID NO:5).

By definition, analogues, homologues and derivatives of HSP70/72 are immunologically related polypeptides. Moreover, all immunologically related polypeptides contain at least one HSP70/72 antigen. Accordingly, "HSP70/72 antigens" may be found in HSP70/72 itself, or in immunologically related polypeptides.

In a further aspect of the invention, we provide polypeptides that are immunologically related to HSP72. As used herein, "immunologically related" polypeptides are characterized by one or more of the following properties:

(a) they are immunologically reactive with antibodies generated by infection of a mammalian host with *Streptococcus pneumoniae* cells, which antibodies are immunologically reactive with HSP72 (SEQ ID NO:5);

(b) they are capable of eliciting antibodies that are immunologically reactive with HSP72 (SEQ ID NO:5);

(c) they are immunologically reactive with antibodies elicited by immunization of a mammal with HSP72 (SEQ ID NO:5).

By definition, analogues, homologues and derivatives of HSP72 are immunologically related polypeptides. Moreover, all immunologically related polypeptides contain at least one HSP72 antigen. Accordingly, "HSP72 antigens" may be found in HSP72 itself, or in immunologically related polypeptides.

As used herein, "related bacteria" are bacteria that possess antigens capable of eliciting antibodies that react with HSP72. Examples of related bacteria include *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus mutans*, *Streptococcus sanguis*, *Streptococcus agalactiae* and *Enterococcus faecalis*.

It will be understood that by following the examples of this invention, one of skill in the art may determine without undue experimentation whether a

particular analogue, homologue, derivative, immunologically related polypeptide, or fragment would be useful in the diagnosis, prevention or treatment of disease. Useful polypeptides and fragments will elicit
5 antibodies that are immunoreactive with HSP72 (Example 4). Preferably, useful polypeptides and fragments will demonstrate the ability to elicit a protective immune response against lethal bacterial infection (Example 5).

Also included are polymeric forms of the
10 polypeptides of this invention. These polymeric forms include, for example, one or more polypeptides that have been crosslinked with crosslinkers such as avidin/biotin, glutaraldehyde or dimethylsuberimidate. Such polymeric forms also include polypeptides containing two or more
15 tandem or inverted contiguous protein sequences, produced from multicistronic mRNAs generated by recombinant DNA technology.

This invention provides substantially pure HSP72 and immunologically related polypeptides. The term
20 "substantially pure" means that the polypeptides according to the invention, and the DNA sequences encoding them, are substantially free from other proteins of bacterial origin. Substantially pure protein preparations may be obtained by a variety of conventional processes, for
25 example the procedures described in Examples 3 and 5.

In another aspect, this invention provides, for the first time, a DNA sequence coding for a heat shock protein of *S. pneumoniae*, specifically, HSP72 (SEQ ID NO:4, nucleotides 682-2502).

30 The DNA sequences of this invention also include DNA sequences coding for polypeptide analogues and homologues of HSP72, DNA sequences coding for immunologically related polypeptides, DNA sequences that are degenerate to any of the foregoing DNA sequences, and
35 fragments of any of the foregoing DNA sequences. It will be readily appreciated that a person of ordinary skill in the art will be able to determine the DNA sequence of any

of the polypeptides of this invention, once the polypeptide has been identified and isolated, using conventional DNA sequencing techniques.

Oligonucleotide primers and other nucleic acid probes derived from the genes encoding the polypeptides of this invention may also be used to isolate and clone other related proteins from *S. pneumoniae* and related bacteria which may contain regions of DNA bacteria that are homologous to the DNA sequences of this invention. In addition, the DNA sequences of this invention may be used in PCR reactions to detect the presence of *S. pneumoniae* or related bacteria in a biological sample.

The polypeptides of this invention may be prepared from a variety of processes, for example by protein fractionation from appropriate cell extracts, using conventional separation techniques such as ion exchange and gel chromatography and electrophoresis, or by the use of recombinant DNA techniques. The use of recombinant DNA techniques is particularly suitable for preparing substantially pure polypeptides according to the invention.

Thus according to a further aspect of the invention, we provide a process for the production of HSP72, immunologically related polypeptides, and fragments thereof, comprising the steps of (1) culturing a unicellular host organism transformed with a vector containing a DNA sequence coding for said polypeptide or fragment and one or more expression control sequences operatively linked to the DNA sequence, and (2) recovering a substantially pure polypeptide or fragment.

As is well known in the art, in order to obtain high expression levels of a transfected gene in a host, the gene must be operatively linked to transcriptional and translational expression control sequences that are functional in the chosen expression host. Preferably, the expression control sequences, and the gene of interest, will be contained in an expression vector that further

comprises a bacterial selection marker and origin of replication. If the expression host is a eukaryotic cell, the expression vector should further comprise an expression marker useful in the eukaryotic expression
5 host.

The DNA sequences encoding the polypeptides of this invention may or may not encode a signal sequence. If the expression host is eukaryotic, it generally is preferred that a signal sequence be encoded so that the
10 mature protein is secreted from the eukaryotic host.

An amino terminal methionine may or may not be present on the expressed polypeptides of this invention. If the terminal methionine is not cleaved by the expression host, it may, if desired, be chemically removed
15 by standard techniques.

A wide variety of expression host/vector combinations may be employed in expressing the DNA sequences of this invention. Useful expression vectors for eukaryotic hosts include, for example, vectors
20 comprising expression control sequences from SV40, bovine papilloma virus, adenovirus, adeno-associated virus, cytomegalovirus, and retroviruses. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, including pBluescript, pGEX2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their
25 derivatives, wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g. λ gt10 and λ gt11, NM989, and other DNA phages, such as M13 and filamentous single stranded DNA phages. Useful
30 expression vectors for yeast cells include the 2 μ plasmid and derivatives thereof. Useful vectors for insect cells include pVL 941.

In addition, any of a wide variety of expression control sequences may be used in these vectors to express
35 the DNA sequences of this invention. Useful expression control sequences include the expression control sequences associated with structural genes of the foregoing

expression vectors. Examples of useful expression control sequences include, for example, the early and late promoters of SV40 or adenovirus, the lac system, the trp system, the TAC or TRC system, the T3 and T7 promoters the major operator and promoter regions of phage lambda, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast alpha-mating system and other constitutive and inducible promoter sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. The T7 RNA polymerase promoter $\Phi 10$ is particularly useful in the expression of HSP72 in *E. coli* (Example 3).

Host cells transformed with the foregoing vectors form a further aspect of this invention. A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of *E. coli*, *Pseudomonas*, *Bacillus*, *Streptomyces*, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO and mouse cells, African green monkey cells such as COS 1, COS 7, BSC 1, BSC 40, and BMT 10, human cells, and plant cells in tissue culture. Preferred host organisms include bacteria such as *E. coli* and *B. subtilis*, and mammalian cells in tissue culture.

It should of course be understood that not all vectors and expression control sequences will function equally well to express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must replicate

in it. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered. In selecting an expression control
5 sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the DNA sequences of this invention, particularly as regards potential secondary structures. Unicellular hosts
10 should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the DNA sequences of this invention, their secretion characteristics, their ability to fold the protein correctly, their fermentation or culture
15 requirements, and the ease of purification from them of the products coded for by the DNA sequences of this invention. Within these parameters, one of skill in the art may select various vector/expression control sequence/host combinations that will express the DNA
20 sequences of this invention on fermentation or in large scale animal culture.

The polypeptides encoded by the DNA sequences of this invention may be isolated from the fermentation or cell culture and purified using any of a variety of
25 conventional methods including: liquid chromatography such as normal or reversed phase, using HPLC, FPLC and the like; affinity chromatography (such as with inorganic ligands or monoclonal antibodies); size exclusion chromatography; immobilized metal chelate chromatography;
30 gel electrophoresis; and the like. One of skill in the art may select the most appropriate isolation and purification techniques without departing from the scope of this invention.

In addition, the polypeptides of this invention
35 may be generated by any of several chemical techniques. For example, they may be prepared using the solid-phase synthetic technique originally described by R. B.

Merrifield, "Solid Phase Peptide Synthesis. I. The Synthesis Of A Tetrapeptide", J. Am. Chem. Soc., 83, pp. 2149-54 (1963), or they may be prepared by synthesis in solution. A summary of peptide synthesis techniques
5 may be found in E. Gross & H. J. Meinhofer, 4 The Peptides: Analysis, Synthesis, Biology; Modern Techniques Of Peptide And Amino Acid Analysis, John Wiley & Sons, (1981) and M. Bodanszky, Principles Of Peptide Synthesis, Springer-Verlag (1984).

10 The preferred compositions and methods of this invention comprise polypeptides having enhanced immunogenicity. Such polypeptides may result when the native forms of the polypeptides or fragments thereof are modified or subjected to treatments to enhance their
15 immunogenic character in the intended recipient. Preferred polypeptides are fragments that are specific to Streptococcal species such as fragments selected from the C-terminal portion of thenative polypeptides. Numerous techniques are available and well known to those of skill
20 in the art which may be used, without undue experimentation, to substantially increase the immunogenicity of the polypeptides herein disclosed. For example, the polypeptides may be modified by coupling to dinitrophenol groups or arsanilic acid, or by denaturation
25 with heat and/or SDS. Particularly if the polypeptides are small polypeptides synthesized chemically, it may be desirable to couple them to an immunogenic carrier. The coupling of course, must not interfere with the ability of either the polypeptide or the carrier to function
30 appropriately. For a review of some general considerations in coupling strategies, see Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, ed. E. Harlow and D. Lane (1988). Useful immunogenic carriers are well known in the art. Examples of such carriers are
35 keyhole limpet hemocyanin (KLH); albumins such as bovine serum albumin (BSA) and ovalbumin, PPD (purified protein derivative of tuberculin); red blood cells; tetanus

toxoid; cholera toxoid; agarose beads; activated carbon; or bentonite.

Modification of the amino acid sequence of the polypeptides disclosed herein in order to alter the lipidation state is also a method which may be used to increase their immunogenicity and biochemical properties. For example, the polypeptides or fragments thereof may be expressed with or without the signal sequences that direct addition of lipid moieties.

In accordance with this invention, derivatives of the polypeptides may be prepared by a variety of methods, including by *in vitro* manipulation of the DNA encoding the native polypeptides and subsequent expression of the modified DNA, by chemical synthesis of derivatized DNA sequences, or by chemical or biological manipulation of expressed amino acid sequences.

For example, derivatives may be produced by substitution of one or more amino acids with a different natural amino acid, an amino acid derivative or non-native amino acid, conservative substitution being preferred, e.g., 3-methylhistidine may be substituted for histidine, 4-hydroxyproline may be substituted for proline, 5-hydroxylysine may be substituted for lysine, and the like.

Causing amino acid substitutions which are less conservative may also result in desired derivatives, e.g., by causing changes in charge, conformation and other biological properties. Such substitutions would include for example, substitution of a hydrophilic residue for a hydrophobic residue, substitution of a cysteine or proline for another residue, substitution of a residue having a small side chain for a residue having a bulky side chain or substitution of a residue having a net positive charge for a residue having a net negative charge. When the result of a given substitution cannot be predicted with certainty, the derivatives may be readily assayed according to the methods disclosed herein to determine the presence or absence of the desired characteristics.

The polypeptides may also be prepared with the objective of increasing stability or rendering the molecules more amenable to purification and preparation. One such technique is to express the polypeptides as fusion proteins comprising other *S. pneumoniae* or non-*S. pneumoniae* sequences. It is preferred that the fusion proteins comprising the polypeptides of this invention be produced at the DNA level, e.g., by constructing a nucleic acid molecule encoding the fusion, transforming host cells with the molecule, inducing the cells to express the fusion protein, and recovering the fusion protein from the cell culture. Alternatively, the fusion proteins may be produced after gene expression according to known methods. An example of a fusion protein according to this invention is the FucI/HSP72 (C-169) protein of Example 3, infra.

The polypeptides of this invention may also be part of larger multimeric molecules which may be produced recombinantly or may be synthesized chemically. Such multimers may also include the polypeptides fused or coupled to moieties other than amino acids, including lipids and carbohydrates.

The polypeptides of this invention are particularly well-suited for the generation of antibodies and for the development of a protective response against disease. Accordingly, in another aspect of this invention, we provide antibodies, or fragments thereof, that are immunologically reactive with HSP72. The antibodies of this invention are either elicited by immunization with HSP72 or an immunologically related polypeptide, or are identified by their reactivity with HSP72 or an immunologically related polypeptide. It should be understood that the antibodies of this invention are not intended to include those antibodies which are normally elicited in an animal upon infection with naturally occurring *S. pneumoniae* and which have not been removed from or altered within the animal in which they were elicited.

The antibodies of this invention may be intact immunoglobulin molecules or fragments thereof that contain an intact antigen binding site, including those fragments known in the art as F(v), Fab, Fab' and F(ab')₂. The antibodies may also be genetically engineered or synthetically produced. The antibody or fragment may be of animal origin, specifically of mammalian origin, and more specifically of murine, rat, monkey or human origin. It may be a natural antibody or fragment, or if desired, a recombinant antibody or fragment. The antibody or antibody fragments may be of polyclonal, or preferably, of monoclonal origin. They may be specific for a number of epitopes but are preferably specific for one. Specifically preferred are the monoclonal antibodies F1-Pn3.1, F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4 of Example 2, infra. One of skill in the art may use the polypeptides of this invention to produce other monoclonal antibodies which could be screened for their ability to confer protection against *S. pneumoniae*, *S. pyogenes*, *S. agalactiae* or other Streptococcal related bacterial infection when used to immunize naive animals. Once a given monoclonal antibody is found to confer protection, the particular epitope that is recognized by that antibody may then be identified. Methods to produce polyclonal and monoclonal antibodies are well known to those of skill in the art. For a review of such methods, see *Antibodies, A Laboratory Manual*, *supra*, and D.E. Yelton, et al., Ann. Rev. of Biochem., 50, pp. 657-80 (1981). Determination of immunoreactivity with a polypeptide of this invention may be made by any of several methods well known in the art, including by immunoblot assay and ELISA.

An antibody of this invention may also be a hybrid molecule formed from immunoglobulin sequences from different species (e.g., mouse and human) or from portions of immunoglobulin light and heavy chain sequences from the same species. It may be a molecule that has multiple binding specificities, such as a bifunctional antibody

prepared by any one of a number of techniques known to those of skill in the art including: the production of hybrid hybridomas; disulfide exchange; chemical cross-linking; addition of peptide linkers between two
5 monoclonal antibodies; the introduction of two sets of immunoglobulin heavy and light chains into a particular cell line; and so forth. The antibodies of this invention may also be human monoclonal antibodies, for example those produced by immortalized human cells, by
10 SCID-hu mice or other non-human animals capable of producing "human" antibodies, or by the expression of cloned human immunoglobulin genes.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety
15 of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way
20 that may render it more suitable for a particular application.

The polypeptides, DNA sequences and antibodies of this invention are useful in prophylactic, therapeutic and diagnostic compositions for preventing, treating and
25 diagnosing disease.

Standard immunological techniques may be employed with the polypeptides and antibodies of this invention in order to use them as immunogens and as vaccines. In particular, any suitable host may be
30 injected with a pharmaceutically effective amount of polypeptide to generate monoclonal or polyvalent antibodies or to induce the development of a protective immunological response against disease. Preferably, the polypeptide is selected from the group consisting of
35 HSP72 (SEQ ID NO:5), HSP70 (SEQ ID NO:20 and SEQ ID NO:22) or fragments thereof.

As used herein, a "pharmaceutically effective amount" of a polypeptide or of an antibody is the amount that, when administered to a patient, elicits an immune response that is effective to prevent or lessen the severity of Streptococcal or related bacterial infections.

The administration of the polypeptides or antibodies of this invention may be accomplished by any of the methods described in Example 10, *infra*, or by a variety of other standard procedures. For a detailed discussion of such techniques, see Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory, ed. E. Harlow and D. Lane (1988). Preferably, if a polypeptide is used, it will be administered with a pharmaceutically acceptable adjuvant, such as complete or incomplete Freund's adjuvant, RIBI (muramyl dipeptides) or ISCOM (immunostimulating complexes). Preferably, the composition will include a water-in-oil emulsion or aluminum hydroxide as adjuvant and will be administered intramuscularly. The vaccine composition may be administered to the patient at one time or over a series of treatments. The most effective mode of administration and dosage regimen will depend upon the level of immunogenicity, the particular composition and/or adjuvant used for treatment, the severity and course of the expected infection, previous therapy, the patient's health status and response to immunization, and the judgment of the treating physician. For example, in an immunocompetent patient, the more highly immunogenic the polypeptide, the lower the dosage and necessary number of immunizations. Similarly, the dosage and necessary treatment time will be lowered if the polypeptide is administered with an adjuvant.

Generally, the dosage will consist of an initial injection, most probably with adjuvant, of about 0.01 to 10 mg, and preferable 0.1 to 1.0 mg, HSP72 antigen per patient, followed most probably by one or maybe more

booster injections. Preferably, boosters will be administered at about 1 and 6 months after the initial injection.

Any of the polypeptides of this invention may be used in the form of a pharmaceutically acceptable salt. Suitable acids and bases which are capable of forming salts with the polypeptides of the present invention are well known to those of skill in the art, and include inorganic and organic acids and bases.

To screen the polypeptides and antibodies of this invention for their ability to confer protection against diseases caused by *S. pneumoniae* or related bacteria, or their ability to lessen the severity of such infection, one of skill in the art will recognize that a number of animal models may be used. Any animal that is susceptible to infection with *S. pneumoniae* or related bacteria may be useful. The Balb/c mice of Example 5, infra, are the preferred animal model for active immunoprotection screening, and the severe-combined immunodeficient mice of Example 5 are the preferred animal model for passive screening. Thus, by administering a particular polypeptide or antibody to these animal models, one of skill in the art may determine without undue experimentation whether that polypeptide or antibody would be useful in the methods and compositions claimed herein.

According to another embodiment of this invention, we describe a method which comprises the steps of treating a patient with a vaccine comprising a pharmaceutically effective amount of any of the polypeptides of this invention in a manner sufficient to prevent or lessen the severity, for some period of time, of Streptococcal or related bacterial infection. Again, the preferred polypeptide for use in such methods is HSP70/HSP72, or fragments thereof.

The polypeptides, DNA sequences and antibodies of this invention may also form the basis for diagnostic methods and kits for the detection of pathogenic

organisms. Several diagnostic methods are possible. For example, this invention provides a method for the detection of *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae* or related bacteria in a biological sample comprising the steps of:

- (a) isolating the biological sample from a patient;
- (b) incubating an antibody of this invention, or fragment thereof with the biological sample to form a mixture; and
- (c) detecting specifically bound antibody or fragment in the mixture which indicates the presence of *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae* or related bacteria. Preferable antibodies for use in this method include monoclonal antibodies F1-Pn3.1, F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4.

Alternatively, this invention provides a method for the detection of antibodies specific to *Streptococcus pneumoniae* or related bacteria in a biological sample comprising:

- (a) isolating the biological sample from a patient;
- (b) incubating a polypeptide of this invention or fragment thereof, with the biological sample to form a mixture; and
- (c) detecting specifically bound polypeptide in the mixture which indicates the presence of antibodies specific to *Streptococcus pneumoniae* or related bacteria. HSP72 (SEQ ID NO:5), the C-169 fragment thereof (residues 439-607 of SEQ ID NO:5), the C-151 fragment thereof (residues 457-607 of SEQ ID NO:5) and peptide fragments GFDAERDAAQAALDD (residues 527-541 of SEQ ID NO:5) and AEGAQATGNAGDDVV (residues 586-600 of SEQ ID NO:5) are the preferred polypeptide and fragments in the above method for the detection of antibodies.

One of skill in the art will recognize that these diagnostic tests may take several forms, including

an enzyme-linked immunosorbent assay (ELISA), a radioimmunoassay or a latex agglutination assay.

The diagnostic agents may be included in a kit which may also comprise instructions for use and other appropriate reagents, preferably a means for detecting when the polypeptide or antibody is bound. For example, the polypeptide or antibody may be labeled with a detection means that allows for the detection of the polypeptide when it is bound to an antibody, or for the detection of the antibody when it is bound to *S. pneumoniae* or related bacteria. The detection means may be a fluorescent labeling agent such as fluorescein isocyanate (FIC), fluorescein isothiocyanate (FITC), and the like, an enzyme, such as horseradish peroxidase (HRP), glucose oxidase or the like, a radioactive element such as ^{125}I or ^{51}Cr that produces gamma ray emissions, or a radioactive element that emits positrons which produce gamma rays upon encounters with electrons present in the test solution, such as ^{11}C , ^{15}O , or ^{13}N . Binding may also be detected by other methods, for example via avidin-biotin complexes. The linking of the detection means is well known in the art. For instance, monoclonal antibody molecules produced by a hybridoma may be metabolically labeled by incorporation of radioisotope-containing amino acids in the culture medium, or polypeptides may be conjugated or coupled to a detection means through activated functional groups.

The DNA sequences of this invention may be used to design DNA probes for use in detecting the presence of *Streptococcus pneumoniae* or related bacteria in a biological sample. The probe-based detection method of this invention comprises the steps of:

- (a) isolating the biological sample from a patient;
- (b) incubating a DNA probe having a DNA sequence of this invention with the biological sample to form a mixture; and

(c) detecting specifically bound DNA probe in the mixture which indicates the presence of *Streptococcus pneumoniae* or related bacteria.

The DNA probes of this invention may also be used for detecting circulating nucleic acids in a sample, for example using a polymerase chain reaction, as a method of diagnosing *Streptococcus pneumoniae* or related bacterial infections. The probes may be synthesized using conventional techniques and may be immobilized on a solid phase, or may be labeled with a detectable label. A preferred DNA probe for this application is an oligomer having a sequence complementary to at least about 6 contiguous nucleotides of HSP72 (SEQ ID NO:4, nucleotides 682-2502).

The polypeptides of this invention may also be used to purify antibodies directed against epitopes present on the protein, for example, using immunoaffinity purification of antibodies on an antigen column.

The antibodies or antibody fragments of this invention may be used to prepare substantially pure proteins according to the invention for example, using immunoaffinity purification of antibodies on an antigen column.

EXAMPLES

In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only, and are not to be construed as limiting the scope of the invention in any manner.

Example 1 describes the identification of HSP72, an immunoreactive heat shock protein according to the invention. Example 2 describes the isolation of monoclonal antibodies against epitopes of HSP72. Example 3 describes the preparation of recombinant HSP72 and fragments of HSP72 according to the invention. Example 4 describes the antigenic specificity and immunoreactivity

of monoclonal antibodies directed against HSP72, and the identification of immunologically related proteins according to the invention. Example 5 describes processes for obtaining substantially pure HSP72, and the use of
5 HSP72 or antibodies against it to protect against experimental *S. pneumoniae* infection. Example 6 describes the preparation of recombinant C-151 fragment of HSP72 according to the invention. Example 7 describes the humoral immune response following the immunization with
10 recombinant HSP72 or fragments of HSP72 according to the invention. Example 8 describes the localization of linear B-cell epitopes on the HSP72. Example 9 describes the *hsp70* genes and HSP70 proteins from *S. agalactiae* and *S. pyogenes*. Example 10 describes the use of HSP72 antigen
15 in a human vaccine.

EXAMPLE 1 - Identification of Immunoreactive
S. pneumoniae Heat Shock Proteins

20 A. Procedures

Unless otherwise noted, the following procedures were used throughout the Examples herein.

25 1. Bacteria

S. pneumoniae strains were provided by the Laboratoire de la Santé Publique du Québec, Sainte-Anne de Bellevue. *S. pneumoniae* strains included type 4 strain 53
30 and type 6 strain 64. If not specified, *S. pneumoniae* type 6 strain 64 was used. Bacterial strains were grown overnight at 37°C in 5% CO₂ on chocolate agar plates.

35 2. Antigen Preparations

Various *S. pneumoniae* antigens were prepared for immunization and immunoassays. Heat-killed whole cell antigens were obtained by incubating bacterial suspensions

in a water bath prewarmed at 56 C for 20 minutes. Detergent-soluble proteins were extracted from *S. pneumoniae* as follows. Heat-killed bacteria were suspended in 10 mM Hepes buffer (4-(2-Hydroxyethyl)-1-piperazinethan-sulfonsäure) (Boehringer Mannheim GmbH, Germany) at pH 7.4 and sonicated at 20,000 Kz/second, four times for 30 seconds. Intact cells and large debris were removed by centrifugation at 1,700 g for 20 minutes. The supernatant was collected and centrifuged at 100,000 g for 60 minutes. The pellet was resuspended in 1 ml of Hepes buffer, and 1 ml of 2% N-lauroyl sarcosine (Sigma Chemical Co., St. Louis, Mo.) was added. The mixture was incubated for 30 minutes at room temperature and the detergent-soluble fraction was harvested by centrifugation at 100,000 g for 60 minutes.

3. Heat Shock Treatment

S. pneumoniae bacteria (type 4, strain 53 and type 6, strain 64) were resuspended in Eagle's Minimal Essential Medium lacking methionine (ICN Biomedicals Inc., Costa Mesa, CA) and supplemented with 1% BIO-X® (Quelab Laboratories, Montreal, Canada) for 15 minutes at 37°C and then divided into fractions of equal volume. The samples were incubated at either 37°C or 45°C for 5 minutes and then labeled with 100 µCi/ml [³⁵S]methionine (ICN) for 10, 30, or 60 minutes at 37°C. The bacteria were harvested and cell extracts were prepared using Tris-HCl lysis buffer as described above, or SDS-PAGE sample buffer.

4. Immunization Of Mice

Female Balb/c mice (Charles River Laboratories, St-Constant, Québec, Canada) were immunized with *S. pneumoniae* antigens. Immune sera to *S. pneumoniae* type 6 strain 64 were obtained from mice immunized, at two-week intervals, by subcutaneous injections of 10⁷ heat-killed bacteria or 20 µg of detergent-soluble pneumococcal

proteins absorbed to aluminum hydroxide adjuvant (Alhydrogel®; Cedarlane Laboratories Ltd., Hornby, Ontario, Canada). Blood samples were collected prior to immunization and at seven days following the first and
5 second immunization.

5. SDS-PAGE and Immunoassays

Cell extracts were prepared for SDS-PAGE, Western blot analysis and radioimmunoprecipitation assay
10 by incubating bacterial suspensions in Tris-HCl lysis buffer (50mM Tris, 150 mM NaCl, 0.1% Na dodecyl sulfate, 0.5% Na deoxycholate, 2% Triton® X-100, 100 µg/ml phenylmethylsulfonylfluoride, and 2µg/ml aprotinin) at pH
15 8.0 for 30 minutes on ice. Lysed cells were cleared by centrifugation and the supernatants were aliquoted and kept frozen at -70 C.

SDS-PAGE were performed on a 10% polyacrylamide gel according to the method of Laemmli [Nature, 227,
20 pp. 680-685 (1970)], using the Mini Protean® system (Bio-Rad Laboratories Ltd., Mississauga, Canada). Samples were denatured by boiling for 5 minutes in sample buffer containing 2% 2-mercaptoethanol. Proteins were resolved by staining the polyacrylamide gel with PhastGel Blue®
25 (Pharmacia Biotech Inc., Baie d'Urfé, Canada). The radiolabeled products were visualized by fluorography. Fluorograms were scanned using a laser densitometer.

Immunoblot procedures were performed according to the method of Towbin et al. [Proc. Natl. Acad. Sci. USA, 76, pp. 4350-4354 (1979)]. The detection of antigens
30 reactive with antibodies was performed by an indirect antibody immunoassay using peroxidase-labeled anti-mouse immunoglobulins and the o-dianisidine color substrate.

Radioimmunoprecipitation assays were performed as described by J.A. Wiley et al. [J. Virol., 66,
35 pp. 5744-5751 (1992)]. Briefly, sera or hybridoma culture supernatants were added to radiolabeled samples containing

equal amounts of [³⁵S]methionine. The mixtures were allowed to incubate for 90 minutes at 4 C with constant agitation. The immune complexes were then precipitated with bovine serum albumin-treated protein A Sepharose (Pharmacia) for 1 hour at 4 C. The beads were pelleted and washed three times in Tris buffered saline at pH 8.0, and the antigen complexes were then dissociated by boiling in sample buffer. The antigens were analyzed by electrophoresis on SDS-PAGE. The gels were fixed, enhanced for fluorography using Amplify® (Amersham Canada Limited, Oakville, Ontario, Canada), dried, and then exposed to X-ray film.

B. Characterization of the Heat Shock Response in *S. pneumoniae*

We studied the heat shock response of *S. pneumoniae* by examining the pattern of protein synthesis before and after a shift from 37°C to 45°C. FIG. 1 shows the results when *S. pneumoniae* type 6 strain 64 (panel A) and type 4 strain 53 (panel B) were grown at 37°C, incubated at 37°C (lanes 1,3,5,7 and 9) or at 45°C (lanes 2, 4, 6, 8 and 10) for 5 minutes, and then labeled with [³⁵S]methionine for 10 minutes (lanes 1,2 and 7,8), 30 minutes (lanes 3,4 and 9,10), or 60 minutes (lanes 5,6).

The fluorogram derived from SDS-PAGE indicated that the synthesis of at least three proteins was increased by increasing the temperature (FIG. 1). The most prominent induced protein was about 72 kDa (HSP72), whereas the other two were approximately 80 kDa (HSP80) and 62 kDa (HSP62). Increased protein synthesis was already apparent after 10 minutes of labeling (FIG. 1, lanes 1, 2 and 7, 8) and became more significant when the labeling period was prolonged to 30 minutes (FIG. 1, lanes 3, 4 and 9, 10) and 60 minutes (FIG. 1, lanes 5, 6). The effect of elevated temperature on the protein synthesis profile of two different *S. pneumoniae* strains

was similar, with HSPs of similar molecular mass being synthesized (compare Panel A (type 6 strain 64) to Panel B (type 4 strain 53) in FIG. 1).

Analysis of the densitometric tracings from scanning the protein synthesis profiles allowed the estimation of the relative amounts of proteins. For example, with respect to heat-shocked *S. pneumoniae* type 6 strain 64, after 10 minutes of labeling, HSP80 and HSP62 made up 2.9% and 6.8% of the labeled proteins, respectively, compared to less than 0.1% at 37°C (FIG. 2). Labeled proteins having an apparent molecular mass of 72 kDa were detected at both 37°C and 45°C conditions (FIG. 2). Radioimmunoprecipitation analysis revealed, however, that HSP72 was undetectable at 37°C (supra; and FIGS. 3, 4 and 6) thus indicating that peak 9 from FIG. 2 corresponds to protein component(s) comigrating with HSP72. Assuming no variation in the labeling of this material, these results would suggest that the amount of HSP72 represents 8.7% of the total labeled cell protein after heat shock treatment. A comparison of the densitometric tracings revealed that cellular proteins corresponding to peaks 4, 10, 13, 17, 19, and 21 were synthesized at almost the same rate irrespective of heat shock treatment (FIG. 2). However, the synthesis of several proteins (peaks 1, 2, 3, 15, 20, 22, 24, and 26) declined considerably in response to heat shock (FIG. 2).

C. Immune Responses to *S. pneumoniae* HSPs

In order to assess the antibody response to pneumococcal HSPs, mouse sera were first assayed by radioimmunoprecipitation. The repertoire of labeled proteins recognized by sera from mice immunized with *S. pneumoniae* antigen preparations are shown in FIGS. 3 and 4. FIG. 3 relates to detergent soluble protein preparations. FIG. 4 relates to heat-killed bacterial preparation. Although many bands were detected by most antisera, HSP72 was a major precipitation product. The

specificity of antibodies for HSP72 was demonstrated by the detection of proteins among heat-shocked products only (FIG. 3, lanes 4, 6, 8 and 10; FIG. 4, lanes 4, 6 and 8). Interestingly, all immunized mice consistently recognized HSP72. The antibodies reactive with the HSP72 were not specific to the strain used during the immunization since strong reactivities were observed with heterologous *S. pneumoniae* HSP72. It should be noted that in addition to HSP72, one sera precipitated comigrating product labeled at both 37°C and 45°C (FIG. 4, lane 4). This 72 kDa-product probably corresponds to component from peak 9 in FIG. 2 and was not detected in immunoblots. HSP62 is another immune target which was precipitated by some but not all immune sera (FIG. 3, lane 6 and, FIG. 4, lanes 4 and 6). None of the sera tested reacted with HSP80. No proteins were precipitated when preimmune sera taken from the mice used in this study were tested for the presence of antibodies reactive with the labeled products.

As depicted in FIGS. 3 and 5, antibodies to HSP72 could be detected after one immunization with either detergent-soluble proteins or whole cells extracts of *S. pneumoniae*. In addition, a marked increase in the antibody response to HSP72 was observed after a second immunization (FIG. 3, compare 4 and 6, and lanes 8 and 10).

The immunoblot patterns of 15 mice immunized with heat-killed *S. pneumoniae* bacteria were remarkably consistent with the results of the previously described radioimmunoprecipitation. Although antibody response variation occurred to a variety of proteins, HSP72 was a major immunoreactive antigen with 8 (53%) positive sera after the first immunization (FIG. 5). Antibodies to HSP72 were detected in 13 out of 15 (87%) immune sera tested after the second immunization. Two other prominent antigens having apparent molecular mass of 53.5 and 47 kDa were detected in 5 (33%) and 7 (47%) sera, respectively

(FIG. 5). The 72 kDa-reactive band was confirmed as the - pneumococcal HSP72 by using recombinant HSP72 antigens (Example 3, infra) in an immunoblot assay. Preimmune sera failed to detect any pneumococcal proteins.

EXAMPLE 2 - Isolation of Monoclonal Antibodies
Against Epitopes of HSP72

A. Procedures

1. Immunization of Mice And Fusion

Female Balb/c mice (Charles River Laboratories) were immunized with *S. pneumoniae* antigens. One set of mice (fusion experiment 1) were immunized by peritoneal injection with 10^7 formalin-killed whole cell antigen from strain MTL suspended in Freund's complete adjuvant, and were boosted at two-week intervals with the same antigen and then with a sonicate from heat-killed bacteria in Freund's incomplete adjuvant. A second group of mice (fusion experiment 2) were immunized three times at three-week intervals with 75 μ g of detergent-soluble pneumococcal antigens extracted from strain 64 (type 6) in 25 μ g of Quil A adjuvant (Cedarlane Laboratories Ltd., Hornby, Ontario, Canada). Three days before fusion, all mice were injected intraperitoneally with the respective antigen suspended in PBS alone. Hybridomas were produced by fusion of spleen cells with nonsecreting SP2/0 myeloma cells as previously described by J. Hamel et al. [J. Med. Microbiol., 23, pp. 163-170 (1987)]. Specific hybridomas were cloned by sequential limiting dilutions, expanded and frozen in liquid nitrogen. The class, subclass, and light-chain type of MABs were determined by ELISA as described by D. Martin et al., [Eur. J. Immunol., 18, pp. 601-606 (1988)] using reagents obtained from Southern Biotechnology Associates Inc. (Birmingham, AL).

2. Subcellular Fractionation

Pneumococci were separated into subcellular
5 fractions according to the technique described by Pearce
et al. [Mol. Microbiol., 9, pp. 1037-1050 (1993)].
Briefly, *S. pneumoniae* strain 64 (type 6) was grown in
Todd Hewitt broth supplemented with 0.5% (w/v) yeast
extract for 6 hours at 37°C and isolated by centrifugation.
10 Cell pellets were resuspended in 25 mM Tris-HCl pH 8.0, 1
mM EDTA, 1 mM phenylmethanesulphonylfluoride (PMSF) and
sonicated for 4 minutes with 15 second bursts. Cellular
debris were removed by centrifugation. The bacterial
membranes and cytoplasmic contents were separated by
15 centrifugation at 98,000 g for 4 hours. The cytoplasmic
(supernatant) and the membrane (pellet) fractions were
adjusted to 1 mg protein per ml and subjected to SDS-PAGE
and immunoblot analyses.

20 B. Identification and Characterization of MAb's to the HSP72 of *S. pneumoniae*

Culture supernatants of hybridomas were
initially screened by dot enzyme immunoassay using whole
25 cells from *S. pneumoniae* strain 65 (type 4) according to
the procedures described in D. Martin et al. (supra).
Positive hybridomas were then retested by immunoblotting
in order to identify the hybridomas secreting MAb's
reactive with the HSP72. Of 26 hybridomas with anti-
30 *S. pneumoniae* reactivity in immunoblot, four were found to
recognize epitopes present on a protein band with an
apparent molecular mass of 72 kDa. The four hybridomas
were designated F1-Pn3.1 (from fusion experiment 1) and
F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4 (from fusion experiment
35 2). Isotype analysis revealed that hybridoma F1-Pn3.1
(from fusion experiment 1) secreted IgG-2ak immunoglobulins,
whereas hybridomas F2-Pn3.2, F2-Pn3.3, and F2-Pn3.4 (from

fusion experiment 2) all secreted IgG_{1k}. The specificity of the MABs for HSP72 was clearly demonstrated by the lack of radioimmunoprecipitation activity against [³⁵S]methionine-labeled *S. pneumoniae* proteins obtained from cultures incubated at 37°C and the immunoprecipitation of a 72kDa-protein with heat shock-derived lysates incubated at 45°C. FIG. 6, (lanes 5 and 6) demonstrates the results obtained for MAb F1-Pn3.1. The same results were obtained with MABs F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4

[³⁵S]methionine-labelled lysates from nonheat-shocked and heat-shocked *S. pneumoniae* cells probed with the MABs were electrophoresed on SDS-PAGE gels and then subjected to Western blot analysis. The resulting immunoblots revealed the presence of HSP72 antigen in both samples. FIG. 7, panel A, shows the results obtained for MAb F1-Pn3.1. The same results were obtained with MABs F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4. Accordingly, the heat shock stress did not significantly increase the reactivity of anti-HSP72 monoclonal antibodies. The fluorograph of the immunoblots, however, clearly showed that the heat shock response had occurred (FIG. 7, panel B). These experiments revealed that the rate of synthesis of *S. pneumoniae* HSP72 increases in response to heat shock, but that the absolute amounts of HSP72 do not increase after heat shock.

C. Cellular localization of HSP72

In order to investigate the cellular location of HSP72, *S. pneumoniae* cell lysates were fractionated by differential centrifugation resulting in a soluble fraction and a particulate fraction, enriched in membrane proteins, supra. Sample containing 15 µg protein of membrane fraction (lane 1) and cytoplasmic fraction (lane 2) of *S. pneumoniae* were electrophoresed on SDS-PAGE, transferred to nitrocellulose and probed with MAB F1-

Pn3.1. In the resulting Western blots, HSP72 was found in both fractions, with the majority of the protein associated with the cytoplasmic fraction (FIG. 8).

5 EXAMPLE 3 - Molecular Cloning, Sequencing
and Expression of Genes Coding
for HSP72 Antigens

10 A. Procedures

1. Strains and Plasmids

Strains and plasmids used in this study are listed in Table 1.

TABLE 1: BACTERIAL STRAINS, PHAGES AND PLASMIDS

Strain, Phage Plasmid	Relevant Characteristics	Reference or Source
<u>E. coli Strains</u>		
JM109	$\Delta(lac-proAB)$ [<i>F'</i> <i>traD proAB</i> <i>lacI^q</i> Z Δ M15]	BRL
Y1090	<i>r_k-m_k-lon supF</i> [pMC9]	Amersham
BL21(DE3)	<i>lacUV5-T7</i> RNA polymerase	Studier et al. (<i>infra</i>)
<u>Phages</u>		
λ gt11	<i>cI857 S100</i> cloning vector	Amersham
λ JBD7	LacZ-HSP72 fusion; 2.3 kb EcoRI fragment in λ gt11	This study
λ JBD17	FucI-HSP72 chimeric; 2.4 kb EcoRI and 2.3 kb EcoRI fragments in λ gt11	This study
<u>Plasmids</u>		
pWSK29	Amp ^r ; low copy number cloning vector	Wang et al. (<i>infra</i>)
pWKS30	same as pWSK29 but opposite multi cloning site	Wang et al. (<i>infra</i>)
pJBD171	same as λ JBD17 but in pWSK29	This study
pJBD177	2.8 kb XhoI-EcoRI fragment in pWKS30 no recombinant HSP72 protein expressed	This study
pJBD179	FucI-HSP72 fusion; 2.4 kb EcoRI and 0.8 kb EcoRI- EcoRV fragments in pWSK29	This study
pT7-5	Amp ^r ; T7 promoter Φ 10	Tabor et al. (<i>infra</i>)
pT7-6	same as pT7-5 but opposite multi cloning site	Tabor et al. (<i>infra</i>)
pJBDF51	same as pJBD179 but in pT7-5	This study
pJBDF62	same as pJBD179 but in pT7-6	This study
pDELTA1	Amp ^r ; Tn 1000	BRL
pJBDA1	same as pJBD179 but in pDELTA1	This study

pJBD291	HSP72; 3.2 kb HindIII fragment in pWSK29	This study
pJBdk51	same as pJBD291 but in pT7-5	This study
pJBDA4	same as pJBD291 but in pDELTA1	This study

E. coli strains were grown in L broth or on L agar at 37°C. When necessary, ampicillin was added to the media at the concentration of 50 µg/ml. Plasmids were isolated by using the Magic/Wizard® Mini-Preps kit (Promega, Fisher Scientific, Ottawa, Canada).

2. General Recombinant DNA Techniques

Restriction endonucleases, T4 DNA ligase, and DNA molecular weight standards were purchased from Boehringer Mannheim Canada, Laval, Quebec or Pharmacia Biotech, Uppsala, Sweden. DNA restriction endonuclease digestion and ligation were performed as described by J. Sambrook et al. [Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory Press, N.Y. (1989)]. Agarose gel electrophoresis of DNA fragments was performed following the procedure of J. Sambrook et al. (supra) using the TAE buffer (0.04 M Tris-acetate; 0.002 M EDTA) from Boehringer Mannheim. DNA fragments were purified from agarose gel by using the Prep-A-Gene® DNA purification kit (Bio-Rad Laboratories Ltd., Mississauga, Ontario). Transformation was carried out by electroporation with the Gene Pulser® (Bio-Rad) following the protocol provided by the manufacturer.

3. Construction and Screening of Genomic Library

A genomic *S. pneumoniae* DNA library was generated in the bacteriophage expression vector λgt11 (λgt11 cloning system, Amersham) according to the

procedure provided by the manufacturer. Chromosomal DNA of *S. pneumoniae* type 6 strain 64 was prepared by following the procedure of J.C. Paton et al. [Infect. Immun., 54, pp. 50-55 (1986)]. The *S. pneumoniae* chromosomal DNA was partially digested with EcoRI, and the 4- to 7-kb fragments were fractionated and purified from agarose gel. The fragments were ligated into λ gt11 arms, packaged, and the resulting phage mixtures used to infect *E. coli* Y1090. Immunoscreening of plaques expressing recombinant HSP72 antigens was performed using HSP72-specific monoclonal antibody F1-Pn3.1, supra. Plaque clones expressing peptides recognized by MAb F1-Pn3.1 were isolated and purified. Liquid lysates were prepared and DNA was purified from a Promega LambdaSorb phage adsorbent according to the manufacturer's directions followed by conventional DNA purification procedures.

4. Southern Blot Analysis

The nonradioactive DIG DNA Labelling and Detection kit, obtained from Boehringer Mannheim, was used to perform Southern blot analysis in this example. The DNA fragments selected for use as probes (infra) were purified by agarose gel electrophoresis and then labelled with digoxigenin (DIG)-11-dUTP. Pneumococcal chromosomal DNA was digested with HindIII and the digests were separated by electrophoresis on an 0.8% SDS-PAGE gel and transformed onto positive charged nylon membranes (Boehringer Mannheim) as described by J. Sambrook et al. (supra). The membrane was then blotted with the DIG-labelled DNA probes according to the protocol of the manufacturer.

5. DNA Sequencing and Sequence Analysis

The DNA fragments sequenced in this example were first cloned into plasmid pDELTA 1 (GIBCO BRL Life

Technologies, Burlington, Ontario). A series of nested deletions were generated from both strands by *in vivo* deletion mediated by Tn 1000 transposon transposition (Deletion Factory System, GIBCO BRL) following the procedures provided by the supplier. These deletions were sized by agarose gel electrophoresis and appropriate deletion derivatives were selected for sequencing by the dideoxynucleotide chain terminating method of F. Sanger et al. [Proc. Natl. Acad. Sci. USA, 74, pp. 5463-5467 (1977)]. To sequence the gaps between deletion templates, oligonucleotides were synthesized by oligonucleotide synthesizer 392 (ABI, Applied Biosystems Inc., Foster City, CA). The sequencing reaction was carried out by PCR (DNA Thermal Cycler 480®, Perkin Elmer) using the Taq DyeDeoxy Terminator Cycle Sequencing kit (ABI), and DNA electrophoresis was performed on automated DNA sequencer 373A (ABI).

6. Expression of Cloned Gene in
E. coli T7 RNA pol/promoter system

High level expression of the cloned gene in this example was achieved by employing the bacteriophage T7 RNA polymerase/promoter system in *E. coli*. The DNA fragment specifying the recombinant protein was ligated into plasmids pT7-5 or pT7-6 [S. Tabor and C.C. Richardson, Proc. Natl. Acad. Sci. USA, 82, PP. 1074-1078 (1985)], in a proper orientation in which the gene to be expressed was placed under the control of phage T7 RNA polymerase specific promoter $\Phi 10$. The resulting plasmid was transformed into *E. coli* strain BL21(DE3) [F.W. Studier, and B.A. Moffatt, J. Mol. Biol., 189, pp. 113-130 (1986)] which carries the T7 RNA polymerase structural gene on its chromosome under the control of the inducible lacUV5 promoter. Upon IPTG induction, the T7 RNA polymerase induced in the BL21(DE3) transformants specifically

transcribed the gene under the control of T7 promoter $\Phi 10$. The overexpressed recombinant proteins were visualized by either Western blotting or Coomassie Blue staining.

5 7. N-terminal Amino Acid Sequence
 Analysis of HSP72

Pneumococcal HSP72 was purified by immunoprecipitation using MAb F1-Pn3.1 (supra) and samples
10 of cell wall extracts of *S. pneumoniae* strain 64 prepared as described by L.S. Daniels et al. [Microb. Pathogen., 1, pp. 519-531 (1986)] as antigen. The immune precipitates were resolved by SDS-PAGE and then transferred to polyvinylidene difluoride (PVDF) membrane by the method of
15 P. Matsudaira [J. Biol. Chem., 262, pp. 10035-10038 (1987)]. PVDF membrane was stained with Coomassie Blue, the HSP72 band excised and then analyzed in an automated protein sequencer (ABI), according to standard procedures.

20 B. Construction of Plasmids Containing
 S. pneumoniae HSP72 Gene Fragments
 Corresponding to C-169

The λ gt11 *S. pneumoniae* genomic DNA library was
25 screened with the HSP72-specific MAb F1-Pn3.1. Seventeen (17) immunoreactive clones were isolated and purified from a total of 1500 phages tested. To confirm the specificity of the proteins expressed by the recombinant phages, Western blot analysis of the recombinant phage lysates was
30 performed. Two groups of clones were identified among the 17 positive clones recognized by MAb F1-Pn3.1 and their representatives were designated as λ JBD7 and λ JBD17 for further characterization. As shown in FIG. 9, whole cell extracts from *S. pneumoniae* strain 64 (lane 1) and phage
35 lysates from *E. coli* infected with λ JBD17 (lanes 2 and 3) or λ JBD7 (lanes 4 and 5) cultured in the presence (+) or absence (-) of IPTG were subjected to 10% polyacrylamide

gel electrophoresis and were electrotransferred to nitrocellulose. The immunoblot was probed with HSP72-specific MAb F1-Pn3.1. Clone λ JBD17 had two EcoRI-EcoRI insert fragments of 2.4 kb and 2.3 kb (FIG. 10), and
5 expressed a chimeric recombinant protein having an apparent molecular mass of 74 kDa on SDS-PAGE gel (FIG. 9, lanes 2 and 3). Clone λ JBD7 was found to contain a 2.3 kb EcoRI insert fragment and produced an apparent fusion protein consisting of LacZ and the 74 kDa chimeric protein
10 expressed from clone λ JBD17. The fusion protein had an apparent molecular mass of 160 kDa as estimated by SDS-PAGE (FIG. 9, lane 5). The expression of the chimeric recombinant protein encoded by phage λ JBD17 was independent of IPTG induction (FIG. 9, lanes 2 and 3)
15 while the expression of the recombinant fusion protein encoded by phage λ JBD7 was dependent on induction of the lac promoter (FIG. 9, lanes 4 and 5).

In an attempt to subclone the HSP72 gene, the pneumococcal DNA insert from clone λ JBD17 was extracted,
20 purified and ligated into a low copy plasmid pWSK29 [R.F. Wang and S.R. Kushner, Gene, 100, pp. 195-199 (1991)] to generate plasmid pJBD171. The insert from pJBD171 was characterized by restriction mapping (Fig. 10B), and a series of subcloning and immunoblotting was carried out to
25 define the boundaries of the gene coding for the antigen reactive with MAb F1-Pn3.1. The region responsible for expression of the 74 kDa chimeric protein was found to localize on the 3.2 kb EcoRI-EcoRV fragment, which consists of the intact 2.4 kb EcoRI-EcoRI fragment and the
30 0.8 kb EcoRI-EcoRV portion of the 2.3 kb EcoRI-EcoRI fragment. The plasmid carrying the 3.2 kb EcoRI-EcoRV insert was designated pJBD179.

C. Expression and DNA Sequence
Analysis of a Chimeric Gene
Coding for C-169

5 To further determine the transcriptional
direction of the gene coding for the 74 kDa chimeric
protein on the 3.2 kb EcoRI-EcoRV fragment, and to
increase the yield of the 74 kDa chimeric protein for
10 immunological study, we decided to express the 74 kDa
chimeric protein in the *E. coli* T7 RNA and T7 promoter
system. The 3.2 kb EcoRI-EcoRV fragment, derived from
pJBD179, was ligated into plasmids pT7-5 and pT7-6 in
which the multi-cloning sites were placed in opposite
15 orientation with respect to the T7 RNA polymerase specific
T7 promoter Φ 10. The ligation mixture was used to
transform *E. coli* JM109 and positive transformants
reactive with MAb F1-Pn3.1 were identified by the colony
lifting method described by J. Sambrook et al. [supra].
20 The resulting recombinant plasmids, derived from pT7-5 and
pT7-6, were designated pJBDF51 and pJBDF62, respectively.
The intact 3.2 kb EcoRI-EcoRV insert in these recombinant
plasmids and their orientation was determined by
restriction mapping. To achieve overexpression of the
25 74 kDa chimeric protein, pJBDF51 and pJBDF62 were
transformed, separately, into *E. coli* BL21(DE3). The
transformants were induced with IPTG (1 mM) for 3 hours at
37°C. The cells were harvested, washed, resuspended in
1% SDS and boiled for 10 minutes. The lysates were then
30 used for SDS-PAGE and immunoblot analysis. As expected,
both transformants produced the 74 kDa chimeric protein
readily detected by Western blotting with MAb F1-Pn3.1
(FIG. 11). However, under the IPTG induction condition,
only transformants BL21(DE3)(pJBDF51) overexpressed the 74
35 kDa chimeric protein (FIG. 11A and B, lane 2) indicating
that the transcriptional direction of the gene on the 3.2

kb EcoRI-EcoRV fragment is from the EcoRI end towards the EcoRV end (FIG. 10A).

The 3.2 kb EcoRI-EcoRV fragment was cloned into plasmid pDELTA 1 to yield plasmid pJBDA1. A series of overlapping deletions were generated and used as DNA sequencing templates. The DNA sequence of the entire 3.2 kb EcoRI-EcoRV insert is SEQ ID NO:1. Two open reading frames ("ORFs") were found and their orientation is indicated in FIG. 10B ("ORF27" and "FucI-HSP72 (C-169)"). In front of these two ORFs, putative ribosome-binding sites were identified (SEQ ID NO:1, nucleotides 18-21 and 760-763). No obvious -10 and -35 promoter sequences were detected. ORF27 spans nucleotides 30-755 (SEQ ID NO:1) and encodes a protein of 242 amino acids with a calculated molecular weight of 27,066 daltons. The deduced amino acid sequence of this protein is SEQ ID NO:2. We designated this gene *orf27*, and compared it to other known sequences. No homologous gene or protein was found. The large ORF (nucleotides 771-2912, SEQ ID NO:1) specifies a protein of 714 amino acids with a predicted molecular mass of 79,238 daltons. The deduced amino acid sequence of this protein is SEQ ID NO:3. This ORF was compared with other known sequences to determine its relationship to other amino acid sequences. This analysis revealed a high degree of similarity of the encoded protein to the sequence of *E. coli* fucose isomerase (FucI) and to several HSP70 gene family members, also known as *DnaK* genes. Alignment of SEQ ID NO:3 and those of the *E. coli* FucI and HSP70 (Dnak) proteins indicated that the N-terminal portion corresponding to amino acids 1 to 545 (SEQ ID NO:3) of the 74 kDa chimeric protein is highly homologous to *E. coli* FucI, while the C-terminal portion corresponding to amino acids 546-714 (SEQ ID NO:3) is similar to HSP70 (Dnak) proteins. It is noteworthy that there is an EcoRI restriction site lying in the junction of these two portions of the gene coding for the 74 kDa protein (SEQ ID NO:1, between nucleotides 2404 and 2405).

Other restriction sites exist between nucleotides 971 and 972 (Pst I), nucleotides 1916 and 1917 (Pst I), nucleotides 1978 and 1979 (Xho I), and nucleotides 3164 and 3165 (EcoRV). From these data we concluded that the 74 kDa protein was a chimeric protein encoded by two pieces of *S. pneumoniae* chromosomal DNA, a 2.4 kb EcoRI-EcoRI fragment derived from the *FucI* homologous gene and a 2.3 kb EcoRI-EcoRI fragment derived from the HSP72 gene.

10 D. Southern Blot Analysis

Southern blotting was performed in order to confirm that the 74 kDa protein is a chimeric protein and to attempt to clone the entire pneumococcal HSP72 gene. Chromosomal *S. pneumoniae* DNA was digested with HindIII to completion, separated on a 0.8% agarose gel, and transferred onto two positively charged nylon membranes (Boehringer Mannheim). The membranes were then blotted with either the 0.8 kb EcoRI-EcoRV probe, derived from the 2.3 kb EcoRI-EcoRI fragment, or the 1 kb PstI-PstI probe, obtained from the 2.4 kb EcoRI-EcoRI fragment. Both probes had been previously labelled with digoxigenin-dUTP. These two probes hybridized two individual HindIII fragments of different sizes (FIGS. 10B and 10C). The 0.8 kb EcoRI-EcoRV probe recognized the 3.2 kb HindIII fragment and the 1 kb PstI-PstI probe reacted with the 4 kb HindIII fragment. This result further indicated that the gene responsible for the expression of the 74 kDa chimeric protein was generated by fusion, in frame, of two pieces of EcoRI fragments, one originated from the fragment containing the 5' portion of the *S. pneumoniae* *FucI* homologue, the other derived from the segment carrying the C-169 fragment of the pneumococcal HSP72 gene. The fact that the 0.8 kb EcoRI-EcoRV probe hybridized a single 3.2 kb fragment suggested that there is only a single HSP72 gene copy in *S. pneumoniae*.

E. Production of Recombinant HSP72

A partial pneumococcal genomic library was generated by ligation of the pool of HindIII digests of chromosomal DNA, with sizes ranging from 2.8 to 3.7 kb, into plasmid pWSK29/HindIII. The ligation mixture was used to transform *E. coli* strain JM 109 and the transformants were screened by hybridization with the 0.8 kb EcoRI-EcoRV probe. One representative plasmid from four positive hybridizing clones was named pJBD291. Restriction analysis of the insert and Western blot of the cell lysate of transformants were employed to verify that the plasmid pJBD291 indeed carries the 3.2 kb HindIII fragment containing the HSP72 gene expressing the recombinant HSP72 protein (FIG. 10B). The HSP72 protein expressed by the transformants (pJBD291) migrated on the SDS-PAGE gel at the same position as the native HSP72 protein (FIG. 12). To sequence the entire HSP72 gene and to overexpress the full-length HSP72 protein, the 3.2 kb HindIII fragment was isolated from plasmid pJBD291, and subcloned into plasmids pDELTA 1 and pT7-5 to generate pJBDA4 and pJBDk51, respectively.

The entire 3.2 kb HindIII DNA fragment carried on the plasmid pJBDA4 and the 2.3 kb EcoRI-EcoRI DNA fragment contained on the plasmid pJBD177 were sequenced. Altogether, the nucleotide sequence comprised 4320 base pairs and revealed two ORFs (SEQ ID NO:4). The first ORF, starting at nucleotide 682 and ending at nucleotide 2502 (SEQ ID NO:4), was identified as the pneumococcal HSP72 gene, and the second ORF, spanning from nucleotide 3265 to nucleotide 4320 (SEQ ID NO:4), was located 764 base pairs downstream from the HSP72 structural gene and was identified as the 5' portion of the pneumococcal *DnaJ* gene. The putative ribosome binding site ("AGGA") was located 9 base pairs upstream from the start codon of the HSP72 structural gene, while the typical ribosome binding

site ("AGGA") was found 66 base pairs upstream from the - start codon of the *DnaJ* structural gene. No typical 5' regulatory region was identified in front of these two genes. Restriction sites are located between nucleotides 5 1 and 2 (HindIII), nucleotides 1318 and 1319 (EcoRI), nucleotides 1994 and 1995 (EcoRI), nucleotides 3343 and 3344 (HindIII), and nucleotides 4315 and 4316 (EcoRI). The gene organization of HSP72 (*DnaK*) and *DnaJ* in *S. pneumoniae* is similar to that of *E. coli* [Saito, H. and 10 Uchida, Mol. Gen. Genet. 164, 1-8 (1978)] as well as several other Gram positive bacteria [Wetzstein, M. et al., J. Bacteriol. 174, 3300-3310 (1992)]. However, the intragenic region of *S. pneumoniae* is significantly larger and no ORF for the *grpE* gene was found upstream of 15 the HSP72 (*DnaK*) structural gene.

The predicted HSP72 protein has 607 amino acids and a calculated molecular mass of 64,755 daltons, as compared to the 72 kDa molecular mass estimated by SDS-PAGE. The predicted HSP72 protein is acidic with an 20 isoelectric point (pI) of 4.35. Automated Edman degradation of the purified native HSP72 protein extracted from *S. pneumoniae* strain 64 revealed SKIIGIDLGTNN-AVAVLE as the 19 amino acid N-terminal sequence of the protein. The amino-terminal methionine was not detected, presumably 25 due to *in situ* processing which is known to occur in many proteins. No amino acid residue was identified on position 13. The 19 amino acid N-terminal sequence obtained from the native HSP72 protein is in full agreement with the 19 amino acid N-terminal sequence 30 deduced from the nucleotide sequence of the recombinant *S. pneumoniae* HSP72 gene (SEQ ID NO:5) thus confirming the cloning. This N-terminal sequence showed complete identity with the *DnaK* protein from *Lactococcus lactis* and 68.4% identity with the *DnaK* protein from *Escherichia* 35 *Coli*. Similarly, the alignment of the predicted amino acid sequence of HSP72 (SEQ ID NO:5) with those from other bacterial HSP70 (*DnaK*) proteins also revealed high

homology (FIGS. 13A-13D). For example, HSP72 showed 54% - identity with the *E. coli* DnaK protein. The highest identity value was obtained from comparison with the Gram positive bacterium *Lactococcus lactis*, showing 85% identity with HSP72. Like other HSP70 proteins of Gram positive bacteria, HSP72 misses a stretch of 24 amino acids near the amino terminus when compared with DnaK proteins from Gram negative bacteria (FIGS. 13A-13D).

Although HSP72 shares homology with HSP70 (DnaK) proteins from other organisms, it does possess some unique features. Sequence divergence of the HSP70 (DnaK) proteins is largely localized to two regions (residues 244 to 330 and 510 to 607, SEQ ID NO:5). More specifically, the peptide sequences GFDAERDAAQAALDD (residues 527 to 541, SEQ ID NO:5) and AEGAQATGNAGDDVV (residues 586 to 600, SEQ ID NO:5) are exclusive to HSP72. The fact that the C-terminal portion of HSP72 is highly variable suggests that this portion carries antigenic determinants specific to *S. pneumoniae*. Consistent with this hypothesis, monoclonal antibodies directed against the C-169 fragment of HSP72 (*infra*), were not reactive with *E. coli* and *S. aureus*, which are known to express DnaK proteins similar to HSP72.

The truncated DnaJ protein of *S. pneumoniae* (SEQ ID NO:6) has 352 amino acids, which show a high degree of similarity with the corresponding portions of the *L. lactis* DnaJ protein (72% identity) and the *E. coli* DnaJ protein (51% identity). The predicted truncated DnaJ protein contains high glycine content (15%). Four Gly-, Cys-rich repeats, each with the Cys-X-X-Cys-X-Gly-X-Gly motif characteristic of DnaJ proteins [P.A. Silver and J.C. Way, *Cell*, 74, pp. 5-6 (1993)], were identified between amino acids 148 and 212 of the *S. pneumoniae* DnaJ protein (SEQ ID NO:6). Three repeated GGFGG sequences (residues 75-79, 81-85, and 90-94) were found near the N-terminus.

F. Reactivity of MABs Against
Recombinant Antigens

5 The four HSP72 specific MABs (F1-Pn3.1, F2-Pn3.2, F2-Pn3.3 and F2-Pn3.4, supra) were tested for their reactivity against proteins expressed by *E. coli* infected or transformed with recombinant phages and plasmids containing HSP72 sequences. The four individual MABs
10 reacted with the lacZ-HSP72 fusion protein expressed by the clone λ JBD7, thus localizing the epitopes recognized by these MABs to the C-terminal 169 residues. Surprisingly, the proteins encoded by the pneumococcal inserts in λ JBD17 and pJBDA1 were recognized by only 3 of
15 4 Mabs. These results suggest that although the C-169 fragments synthesized in *E. coli* infected with λ JBD7 and λ JBD17 have the same primary structure, they have distinct conformation. The lack of reactivity of MAB F2-Pn3.2 with some recombinant proteins raised the possibility that this
20 particular MAB recognizes a more complex epitope. Although complex, F2-Pn3.2 epitopes are still recognizable on Western immunoblots. The complete HSP72_{rec} protein expressed by *E. coli* containing the recombinant plasmid pJBDA4 was reactive with all four MABs.

25 EXAMPLE 4 - Antigenic Specificity and
 Reactivity of HSP72-Specific
 Monoclonal Antibodies

30 The reactivity of MABs F1-Pn3.1, F2-3.2., F2-Pn3.3 and F2-Pn3.4 to a collection of bacterial strains including 20 *S. pneumoniae* strains representing 16 capsular serotypes (types 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 19, 20, and 22) and the 17 non-pneumococcal
35 bacterial strains listed in Table 2, was tested using a dot enzyme immunoassay as described by D. Martin et al. [supra] and immunoblotting. For dot enzyme immunoassay, the bacteria were grown overnight on chocolate agar plates

and then suspended in PBS, pH 7.4. A volume of 5 µl of a suspension containing approximately 10^9 CFU/ml was applied to a nitrocellulose paper, blocked with PBS containing 3% bovine serum albumin, and then incubated sequentially with
 5 MAbs and peroxidase-labeled secondary antibody. Whole cell extracts were prepared for Western blot analysis by boiling bacterial suspensions in sample buffer for 5 minutes.

**TABLE 2:LIST OF NON-PNEUMOCOCCAL ISOLATES
TESTED BY DOT ENZYME IMMUNOASSAY**

<u>Strain</u> <u>Designation</u>	<u>Genus species</u>	<u>group or type</u>
C-2	<i>Streptococcus pyogenes</i>	group A
C-3	<i>Streptococcus agalactiae</i>	group B
C-7	<i>Enterococcus faecalis</i>	group D
C-9	<i>Streptococcus bovis</i>	group D
C-14	<i>Streptococcus mutans</i>	
C-15	<i>Streptococcus salivarius</i>	
C-19	<i>Streptococcus sanguis</i>	I
C-20	<i>Streptococcus sanguis</i>	I
C-21	<i>Streptococcus sanguis</i>	I
C-22	<i>Streptococcus sanguis</i>	II
C-23	<i>Streptococcus sanguis</i>	II
C-24	<i>Streptococcus sanguis</i>	II
C-25	<i>Streptococcus sanguis</i>	II
C-27	<i>Gemella morbillorum</i>	
C-30	<i>Staphylococcus aureus</i>	
C-33	<i>Bacillus</i>	
C-36	<i>Escherichia coli</i>	

10

When tested by dot enzyme immunoassay, each MAb reacted with each of the *S. pneumoniae* strains and none of the non-pneumococcal isolates. These results were unexpected since comparison studies revealed that HSP72 is

very similar to other known bacterial HSP70 (DnaK) proteins, for example those from *E. coli* and *S. aureus*.

Immunoblots were then performed to further investigate the immunoreactivities of our MABs. As shown in Table 3, each MAB exhibited some reactivity. Although the percent identity of the *E. coli* amino acid sequence and the HSP72 amino acid sequence (SEQ ID NO:5) is 54%, the four HSP72-specific MABs did not recognize the *E. coli* HSP70 (DnaK) protein. Similarly, the HSP72-specific MABs did not react with the *C. trachomatis* HSP70 (DnaK) protein, which has 56% amino acid identity with the amino acid sequence of HSP72. High amino acid sequence homology is observed between HSP72 and the HSP70 (DnaK) proteins from gram positive bacterial species. However, again, none of the HSP72-specific MABs reacted with *S. aureus* or *Bacillus* gram positive species, which exhibit 74% and 76% amino acid sequence homology, respectively, with HSP72. From these data it is clear that although HSP70 (DnaK) proteins may be structurally related to HSP72, they are immunologically distinct. Among the non-pneumococcal isolates that reacted with at least one MAB, there is *S. pyogenes*, *Enterococcus faecalis*, *S. mutans* and *S. sanguis*, which all belong to the *Streptococcus* or *Streptococcus*-related *Enterococcus* genus. So far, neither the HSP70 protein, nor the gene structure has been identified in these *Streptococcus* or *Enterococcus* species. Altogether, these observations indicate that hypervariable amino acid sequences or residues within HSP70 (DnaK) proteins are involved in antigenicity. Interestingly, immunoblotting analysis revealed that there was no significant variation in the molecular mass of the HSP70 (DnaK) proteins among both *S. pneumoniae* isolates and immunoreactive non-pneumococcal isolates.

TABLE 3: REACTIVITY OF MABS WITH NON-PNEUMOCOCCAL ISOLATES IN WESTERN IMMUNOBLOTTING

<u>Bacterial Strain</u>			<u>MABs</u>			
<u>Designation</u>	<u>genus/species</u>	<u>type</u>	<u>F1- PN3.1</u>	<u>F2- Pn3.2</u>	<u>F2- PN3.3</u>	<u>F2- Pn3.4</u>
C-2	<i>Streptococcus pyogenes</i>	group A	-	+	-	± ^a
C3	<i>Streptococcus agalactiae</i>	group B	-	-	-	-
C-7	<i>Enterococcus faecalis</i>	group D	-	+	-	-
C-9	<i>Streptococcus bovis</i>	group D	-	-	-	-
C-14	<i>Streptococcus mutans</i>		-	+	-	±
C-15	<i>Streptococcus salivarius</i>		-	-	-	-
C-19	<i>Streptococcus sanguis</i>	I	+	+	-	-
C-20	<i>Streptococcus sanguis</i>	I	+	+	-	+
C-21	<i>Streptococcus sanguis</i>	I	+	+	+	+
C-22	<i>Streptococcus sanguis</i>	II	+	+	+	+
C-23	<i>Streptococcus sanguis</i>	II	+	+	-	-
C-24	<i>Streptococcus sanguis</i>	II	+	+	+	+
C-25	<i>Streptococcus sanguis</i>	II	+	+	+	+
C-27	<i>Gemella morbillorum</i>		-	-	-	-
C-30	<i>Staphylococcus aureus</i>		-	-	-	-
C-33	<i>Bacillus</i>		-	-	-	-
C-36	<i>Escherichia coli</i>		-	-	-	-
C-RP	<i>Chlamydia trachomatis</i> ^b	L2	-	-	-	-

^a ± indicates a weak signal compared to the reactivity observed with *S. pneumoniae* antigens

^b *C. trachomatis* purified elementary bodies were tested.

EXAMPLE 5 - Purification of HSP72 And Its
Use As An Immunogen to Protect
Against Lethal *S. Pneumoniae* Infection

A. Procedures

1. Preparation of Purified
Recombinant HSP72 Protein
and Recombinant C-169

High level exclusive expression of the HSP72 gene was achieved by employing the bacteriophage T7 RNA polymerase/T7 promoter system in *E. coli*. The 3.2 kb HindIII fragment was cloned in both orientations in front of the T7 promoter Φ 10 in the plasmid pT7-5. The resulting plasmid pJBDk51 was then transformed into *E. coli* strain BL21 (DE3). Overexpression of the recombinant HSP72 protein (HSP72_{rec}) was induced by culturing in broth supplemented with antibiotics for a 3-hour period after the addition of IPTG to a final concentration of 1 mM. *E. coli* expressing high levels of HSP72_{rec} were concentrated by centrifugation and lysed by mild sonication in 50 mM Tris-Cl (pH 8.0), 1 mM EDTA and 100 mM NaCl lysis buffer containing 0.2 mg/ml lysozyme. The cell lysates were centrifuged at 12,000 g for 15 minutes and the supernatants were collected. HSP72_{rec} was purified by immunoaffinity using monoclonal antibody F1-Pn3.1 immobilized on sepharose 4B beads (Pharmacia). The purity of eluates was assessed on SDS-PAGE.

The recombinant C-169 protein (C-169_{rec}) was expressed in the form of insoluble inclusion bodies in *E. coli* strain JM109 transformed with the plasmid pJBDA1. Protein inclusion bodies were recovered from pelleted bacterial cells disrupted by sonication as described before. The pellets were washed in lysis buffer containing 1 mg/ml of deoxycholate to remove contaminating materials, and the protein inclusion bodies were then solubilized in urea 6 M. The protein solution was

centrifuged at 100,000 g and the cleared supernatant collected and dialysed against phosphate-buffered saline. After purification, the protein content was determined by the Bio-Rad protein assay (Bio-Rad Laboratories, Mississauga, Ontario, Canada).

2. Active Immunoprotection Studies

Two groups of 10 female Balb/c mice (Charles River Laboratories) were immunized subcutaneously three times at two-week intervals with 0.1 ml of purified HSP72_{rec} or C-169_{rec} antigens adsorbed to Alhydrogel adjuvant. Two antigen doses, approximately 1 and 5 µg, were tested. A third group of 10 control mice were immunized identically via the same route with Alhydrogel adjuvant alone. Blood samples were collected from the orbital sinus prior to each immunization and five to seven days following the third injection. The mice were then challenged with approximately 10⁶ CFU of the type 3 *S. pneumoniae* strain WU2. Samples of the *S. pneumoniae* challenge inoculum were plated on chocolate agar plates to determine the CFU and to verify the challenge dose. Deaths were recorded at 6-hour intervals for the first 3-4 days post-infection and then at 24-hour intervals for a period of 14 days. On days 14 or 15, the surviving mice were sacrificed and blood samples tested for the presence of *S. pneumoniae* organisms. Antibody responses to the recombinant HSP72 antigens are described in Example 7.

3. Passive Immunoprotection Studies

One NZW rabbit (Charles River Laboratories) was immunized subcutaneously at multiple sites with approximately 50 µg of the purified C-169_{rec} protein adsorbed to Alhydrogel adjuvant. The rabbit was boosted three times at two-week intervals with the same antigen and blood samples collected 7 and 14 days following the

last immunization. The serum samples were pooled and antibodies were purified by precipitation using 40% saturated ammonium sulfate.

Severe-combined immunodeficient SCID mice were injected intraperitoneally with 0.25 ml of the purified rabbit antibodies 1 hour before intravenous challenge with 5000 or 880 CFU of the type 3 *S. pneumoniae* strain WU2. Control SCID mice received sterile buffer or antibodies purified from nonimmune rabbit sera. Samples of the *S. pneumoniae* challenge inoculum were plated on chocolate agar plates to determine the CFU and to verify the challenge dose. The SCID mice were chosen because of their high susceptibility to *S. pneumoniae* infection. Blood samples (20 µl each) obtained 24 hours post-challenge were plated on chocolate agar and tested for the presence of *S. pneumoniae* organisms. The level of detection was 50 CFU/ml. Deaths were recorded at 24-hour intervals for a period of 5 days.

B. Results

The availability of cloned *S. pneumoniae* DNA inserts encoding the complete or partial (C-169) HSP72 protein and the expression of recombinant proteins in *E. coli* allowed the obtention of purified proteins useful for the investigation of the vaccinogenic potential of HSP72 protein. Both HSP72_{rec} and C-169_{rec} proteins were obtained in a relatively pure state with no contaminants detected on Coomassie Blue-stained SDS polyacrylamide gels (FIGS. 14 and 15, respectively).

To evaluate the vaccinogenic potential of HSP72, we first examined the ability of HSP72_{rec} to elicit a protective immune response. Groups of 10 mice were immunized with full-length HSP72_{rec} (1 µg or 5 µg dose) and challenged with 4.2 million CFU of *S. pneumoniae* type 3 strain WU2. Eighty percent (80%) of the mice dosed with 1 µg HSP72_{rec} survived the challenge, as did 50% of the mice

dosed with 5 µg HSP72. None of the naive mice immunized with Alhydrogel adjuvant alone without antigen survived the challenge (FIG. 16). No *S. pneumoniae* organisms were detected in any of the blood samples collected on days 14 or 15 from mice surviving infection. The observation that HSP72_{rec} elicited protection against type 3 strain WU2 pneumococci indicated that HSP72 derived from DNA extracted from a type 6 strain contains epitopes capable of eliciting protection against a heterologous strain having a different capsular type.

We further examined the immune response to the HSP72 protein by using recombinant protein fragments expressed from *E. coli* transformed with a chimeric *fucI-HSP72* gene. Mice immunized with purified C-169_{rec} were protected from fatal pneumococcal challenge, thus demonstrating that some, if not all, epitopes eliciting protection are present in the C-terminal region of the HSP72 molecule comprising the last 169 residues. Groups of 10 mice were immunized with C-169_{rec} (1 µg or 5 µg doses) and challenged with 6 million CFU of *S. pneumoniae* type 3 strain WU2. Sixty percent (60%) of the mice dosed with 1 µg C-169_{rec} survived the challenge, as did 70% of the mice dosed with 5 µg C-169_{rec} (FIG. 17). In contrast, all of the naive mice were dead by 2 days post-challenge. Therefore, the C-terminal portion of *S. pneumoniae* HSP72, which includes the region of maximum divergence among DnaK proteins, is a target for the protective immune response.

As illustrated in Table 4 below, two independent experiments demonstrated that SCID mice passively transferred with rabbit anti-C-169_{rec} antibodies were protected from fatal infection with *S. pneumoniae* WU2. In contrast, none of the 15 control mice survived. The control mice received antibodies from nonimmune rabbit sera or received sterile buffer alone. In addition, all mice from the control groups had positive *S. pneumoniae* hemoculture 24 hours post-challenge, while *S. pneumoniae*

organisms were detected in only 2 out of a total of 10 immunized SCID mice.

TABLE 4: PASSIVE IMMUNIZATION STUDIES SHOWING PROTECTION OF SCID MICE FROM EXPERIMENTAL *S. PNEUMONIAE* INFECTION BY ANTI-C-169_{rec} RABBIT ANTIBODIES

<u>Experiment</u>	<u>Injection</u>	<u>No. of Mice Surviving Challenge after 5 days</u>	<u>No. of Mice Testing Positive for the Presence of <i>S. pneumoniae</i></u>
1	sterile	0/5	5/5
	buffer		
	anti-C-169 _{rec}	4/5	2/5
2	control	0/5	5/5
	antibodies		
	sterile	0/5	5/5
	buffer		
	anti-C-169 _{rec}	5/5	0/5

5

In experiments 1 and 2 (Table 4), mice were challenged with 5000 and 880 CFU of type 3 *S. pneumoniae* strain WU2, respectively. Results in Table 4 are expressed as the number of mice surviving challenge, or testing positive for the presence of *S. pneumoniae*, compared to the total number of mice in each group.

Demonstration of the anti-HSP72 specificity of the antibody elicited by immunization with recombinant HSP72 or C-169 proteins came from Western Blot analyses using *S. pneumoniae* cell lysates as antigens. A single band corresponding to HSP72 was detected by all rabbit and mouse antisera tested. These serologic results suggested that the protection following the immunization with recombinant proteins was due to the production of antibodies reactive with *S. pneumoniae* HSP72.

EXAMPLE 6 - Heat-Inducible Expression System for High Level Production of the C-151 Terminal Portion of the HSP72 Protein

A. Construction of Plasmid pURV3 Containing the C-151 terminal coding region of the HSP72 of *S. pneumoniae*

5 The DNA region coding for 151 amino acids at the carboxyl end of the HSP72 of *S. pneumoniae* was inserted downstream of the promoter λ PL into the translation vector p629 [H. J. George et al., Bio/Technology 5, pp. 600-603 (1987)]. This vector
10 contains a cassette of the bacteriophage λ cI857 temperature sensitive repressor gene from which the functional P_R promoter has been deleted. The inactivation of the cI857 repressor by a temperature increase from the ranges of 30-37°C to 37-42°C results in the induction of
15 the gene under the control of λ PL. The induction of gene expression in *E. coli* cells by a temperature shift is advantageous for large scale fermentation since it can easily be achieved with modern fermenters. However, it should be understood that while *E. coli* was the
20 microorganism of choice in the experiments herein described, other host organisms, such as yeast, are intended to be included within the scope of this invention.

 A fragment of 477 nucleotides, including the
25 region of 457 bases between 2050 to 2506 in HSP72 gene of *S. pneumoniae* (see SEQ ID NO 4), was amplified by the polymerase chain reaction (PCR) from the *S. pneumoniae* type 6 strain 64 genomic DNA using the oligonucleotide primers OCRR26 (5'-GGCAGATCTATGAAGGCCAAAGACCTTGGAAC)
30 and OCRR27 (5'-CGCGGATCCTTACTTTTCCGTAAACTCTCCGT). Chromosomal DNA was prepared from a 90 ml culture of exponentially growing cells of *S. pneumoniae* in heart infusion broth using the method of Jayarao et al. [J. Clin. Microbiol., 29, pp. 2774-2778 (1991)]. DNA
35 amplification reactions were made using a DNA Thermal Cycler, Perkin Elmer, San Jose, CA. In OCRR26, an ATG start codon is present in frame just upstream of the

coding region for the amino-terminus region of the C-151. The primers OCRR26 and OCRR27 contain, respectively, a *Bgl*III (AGATCT) and a *Bam*HI (GGATCC) recognition site in order to facilitate the cloning of the PCR product into the dephosphorylated restriction sites *Bgl*III and *Bam*HI of p629. The PCR product was purified from agarose gels by the method of phenol freeze [S. A. Benson, *Biotechniques* 2, pp. 67-68 (1984)] and digested with the restriction enzymes *Bgl*III and *Bam*HI. The *Bgl*III-*Bam*HI fragment of 471 base pairs was then ligated into the *Bgl*III and *Bam*HI recognition sites dephosphorylated of p629. A partial map of the resulting plasmid pURV3 is shown in FIG. 18. This plasmid was transformed by the method of Simanis [Hanahan, D. In D. M. Glover (ed.), *DNA Cloning*, pp. 109-135, (1985)] into the *E. coli* strain XLI Blue MRF' ($\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 \text{ endA1 supE44 thi-1 recA1 gyrA96 relA1 lac}$ [F' *proAB lacI^qZAM15 Tn10 (Tet^r)*]^C) which was obtained from Stratagene, La Jolla, CA. The transformants grown at 37°C were screened by colony immunoblot [J. Sambrook et al. (*supra*)] using the MAb F1-Pn3.1 reactive with C-169_{rec}. Plasmid DNA was purified from a selected transformant and the DNA insert was sequenced by PCR using the Taq Dye Deoxy Terminator Cycle Sequencing kit of Applied Biosystems Inc. (ABI) and DNA electrophoresis was performed on automated DNA sequencer 373A (ABI). The nucleotide sequence of the insert perfectly matched the nucleotide sequence of the C-151 coding region of the HSP72 gene. (See SEQ ID No: 25 and corresponding amino acid sequence at SEQ ID No: 26.) The plasmid was transformed into the prototrophic *E. coli* strain W3110 (ATCC 27325) for the production of C-151_{rec}.

B. Expression of C-151_{rec} and Antigen Preparation

The recombinant C-151_{rec} was synthesized with a methionine residue at its amino end in *E. coli* strain W3110 harboring the plasmid pURV3. *E. coli* cells were

grown at 30°C in LB broth containing 100 µg of ampicillin-per ml until the A₆₀₀ reached a value of 0.6. The cells were then cultivated at 40°C for 18 hours to induce the production of C-151_{rec} protein. A semi-purified C-151_{rec} protein was prepared using the following procedures. The bacterial cells were harvested by centrifugation and the resulting pellet was washed and resuspended in phosphate-buffered saline. Lysozyme was added and the cells were incubated for 15 min on ice before disruption by pulse sonication. The cell lysates were cleared by centrifugation and the supernatants were collected and subjected to separation using an Amicon's ultrafiltration equipment (stirred cells series 8000, Amicon Canada Ltd. Oakville, Ontario). The ultrafiltrate not retained by a YM30 membrane was recovered, analysed by SDS-PAGE and stained with Coomassie blue R-250. Protein concentrations were estimated by comparing the staining intensity of the C-151_{rec} protein with those obtained with defined concentrations of soybean trypsin inhibitor.

20

C. Reactivity of MAbs Against C-151_{rec}

A panel of 10 monoclonal antibodies selected for their reactivity with the *S. pneumoniae* HSP72 protein were tested for their reactivity to C-151_{rec} by Western blot analysis using YM30-ultrafiltrates prepared as described above. The MAbs included a series of six monoclonal antibodies raised to the HSP72_{rec} protein (F3-Pn3.5 to F3-Pn3.10) and monoclonal antibodies F1-Pn3.1, F2-Pn3.2, F2-Pn3.3, F2-Pn3.4. The three MAbs F1-Pn3.1, F2-Pn3.3 and F2-Pn3.4 that were reactive with C-169_{rec} also recognized the C-151_{rec} fragment. All other MAbs were only reactive with HSP72_{rec} thus indicating that they may be directed against epitopes present in the amino terminal region of the HSP72 protein.

35

EXAMPLE 7 - Antibody Response of Balb/c Mice and *Macaca-Fascicularis* (*cynomolgus*) Monkeys to Recombinant HSP72 Antigens

A. Procedures

5 1. Immunization of Animals

Groups of 10 female Balb/c mice were immunized subcutaneously with either HSP72 rec or C-169 rec as described in Example 5. In order to assess the antibody response to C-151rec, a group of 6 mice were immunized 10 three times at two-week intervals with 0.5 µg of C-151rec absorbed to Alhydrogel adjuvant by intraperitoneal injection. Sera from blood samples collected prior each immunization and four to seven days after the third immunization were tested for antibody reactive with *S.* 15 *pneumoniae* by ELISA using plates coated with *S. pneumoniae* cell wall extracts.

Female cynomolgus monkeys were immunized intramuscularly at Day 1, 22 and 77 with 0.5 ml containing 150 µg of purified HSP72rec or C-169rec antigens absorbed 20 to Alhydrogel adjuvant. Blood samples were collected regularly before and after each immunization and the sera were tested for antibody reactive with *S. pneumoniae* HSP72 antigen by Western blot analysis.

The specificity of the raised antibodies for *S.* 25 *pneumoniae* HSP72 was confirmed by Western blot analyses to *S. pneumoniae* cell extracts and purified recombinant antigens.

B. Results

30 The results previously described in Example 5 clearly demonstrate the protective nature of the antibody response elicited following immunization with recombinant HSP72 antigens. Here we monitored the appearance of serum antibody response in mice (FIG. 19, 20 and 21) and in 35 monkeys (FIG. 22) during the immunization schedule. Both species responded strongly to the full-length and truncated recombinant HSP72 proteins used as immunogens

with average titers of 1:64000 after the third injection.- Detailed analysis of individual sera revealed that each animal responded to the immunization in developing antibodies reactive with *S. pneumoniae* HSP72.

5 In mice immunized with C-169_{rec}, the two doses tested, i.e. 1 and 5 µg, were similarly efficient with the induction of similar antibody titers (FIG. 20). A strong boost response was observed after the second injection with C-169_{rec} with no enhancement in the antibody titers
10 after a third injection. In contrast to this, we observed that the immune response to the HSP72_{rec} was dose-dependent. Increases in the specific antibody titers were observed after a second and a third injection with either HSP72_{rec} or C-151_{rec} (FIG. 19 and 21).

15 Study of the immune response of monkeys clearly indicated that the immunogenicity of recombinant HSP72 antigens is not restricted to rodents such as rabbit and mouse. The humoral response following the second injection with either antigen is characterized by a strong
20 increase in HSP72-specific antibody titers that can persist for several weeks without any detectable decrease in their antibody titers (FIG. 22). In addition, specific serum antibodies were detectable in the sera of each monkey after a single injection of recombinant antigens.

25
EXAMPLE 8 - B-Cell Epitope Mapping of HSP72 Stress Protein

 In Example 3, it was shown that significant
30 variability in the primary sequence of the HSP70 proteins was mainly localized to two regions corresponding to amino acid residues 244 to 330 and 510 to 607 of the *S. pneumoniae* HSP72 protein. These variable regions may contain B-cell epitopes responsible for the antigenic
35 heterogeneity reported in Example 4. To investigate this possibility, the reactivity of polyclonal and monoclonal

antibodies to *S. pneumoniae* HSP72 were tested against fourteen peptides selected to cover most of these regions.

A. Procedures

Fourteen peptides of 14 to 30 amino acids residues were synthesized. The peptide sequences and their locations in the protein are summarized in Table 5. Peptides CS870, CS873, CS874, CS875, CS876, CS877, CS878, CS879, CS880 and CS882 were synthesized by Biochem Immunosystem Inc. (Montreal, Canada) using an automated peptide synthesizer. Peptides MAP1, MAP2, MAP3 and MAP4 were synthesized onto a branching lysine core as Multiple Antigenic Peptides (MAP) by the Service de Séquence de Peptides de l'Est du Québec, Centre de recherche du CHUL (Sainte-Foy, Canada). Peptides were purified by reverse-phase high-pressure liquid chromatography. Peptides were solubilized in distilled water except for peptides CS874 and CS876 which were solubilized in a small volume of either 6M guanidine-HCl or dimethyl sulfoxide and then adjusted to 1 mg/ml with distilled water.

Peptide ELISA were performed by coating synthetic peptides onto Immunolon 4 microtitration plates (Dynatech Laboratories, Inc., Chantilly, VA) at a concentration of 50 µg/ml according to the procedures described in J. Hamel et al. [supra]. To confirm the reactivity of MABs with peptides, the ability of fluid-phase peptides to inhibit MAB binding to solid HSP72 was determined. For the inhibition assay, microtitration plates were coated with *S. pneumoniae* cell wall extracts. Hybridoma culture supernatants containing the HSP72-specific MABs were incubated overnight at 4°C with several concentrations of peptide. Peptide treated and control supernatants were then tested by ELISA as described above.

Immune sera were from animals immunized three times with recombinant HSP72 antigens. One rabbit was immunized with 37.5 µg of purified HSP72_{rec} according to the immunization protocol described in Example 5. Pool murine sera were from three Balb/c mice immunized with

HSP72_{rec} from Example 5 and monkey pool sera were from groups of two animals immunized with either HSP72_{rec} or C-169_{rec}.

5 **TABLE 5: SEQUENCES AND LOCATIONS OF SYNTHETIC PEPTIDES CORRESPONDING TO S. PNEUMONIAE HSP72 AMINO ACID RESIDUES**

Peptide	Location	Sequence	Sequence ID No.
CS876	247-261	TSTQISLPFITAGEA	7
CS877	257-271	TAGEAGPLHLEMTLT	8
CS878	268-281	MTLTRAKFDDLTRD	9
CS879	276-290	DDLTRDLVERTKVPV	10
CS880	286-299	TKVPVRQALSDAGL	11
CS882	315-333	RIPAVVEAVKAETGKEPNK	23
CS873	457-471	KAKDLGTQKEQTIVI	12
CS874	467-481	QTIVIQSNSGLTDEE	24
CS875	477-491	LTDEIDRMMKDAAEA	13
MAP 1	487-510	KDAEANAESDKKRKEEVDLRNEVD	14
CS870	507-521	NEVDQAIFATEKTIK	15
MAP 2	517-544	EKTIKETEGKGFDAERDAAQAALD DLKK	16
MAP 3	544-573	KAQEDNNLDDMKAKLEALNEKAQG LAVKLY	17
MAP 4	583-607	QEGAEGAQATGNAGDDVVDGEFTE K	18

10 **B. Identification and Localization of Linear B-Cell Epitopes**

The results presented in FIG. 23 revealed that most of the immunological reactivity was observed with the

peptides localized within amino acid residues 457 and 607 corresponding to the C-151 fragment of HSP72. Rabbit, mice and monkey sera antibody from animals immunized with either recombinant HSP72_{rec} or C-169_{rec} were reactive with both, peptide MAP2 and peptide MAP4. Interestingly, the sequence of peptides MAP2 and MAP4 spans the hypervariable carboxyl-terminal region containing the sequences GFDAERDAAQAALDD (residues 527 to 541) and AEGAQATGNAGDDVV (residues 586 to 600) defined as exclusive to *S. pneumoniae* HSP72 based on the comparison of HSP70 protein sequences available in the data banks. Our data thus revealed that both peptide sequences contain linear B-cell epitopes. In addition, the peptide MAP4 alone was also recognized by the MAbs F1-Pn3.1. This reactivity was confirmed by fluid-phase inhibition assays in which 10 µg/ml of MAP4 caused complete inhibition of F1-Pn3.1 binding to HSP72. Polyclonal antisera from animals immunized with the complete HSP72 recombinant protein also recognized B-cell epitopes localized on peptides CS875, MAP1 and MAP3. All together these data indicate that the hypervariable C-151 terminal fragment of the HSP72 stimulates B-cell responses and possibly constitutes the immunodominant portion of the HSP72 protein. The lack of reactivity of MAbs F2-Pn3.3 and F2-Pn3.4 with the synthetic peptides suggest that they react with conformational determinants present on the C-terminal region of the HSP72. The existence of protective epitopes in the C-151 region was strongly suggested in Example 5 where mice immunized with purified C-169_{rec} were protected from fatal infection with a virulent strain of *S. pneumoniae* thus suggesting that the carboxyl-terminal fragments C-169 or C-151 of *S. pneumoniae* HSP72 or even smaller fragments thereof may prove very useful for the development of a future vaccine.

The variable region comprised within the amino acid residues 244 to 330 also constitutes an antigenic domain. Linear epitopes located on overlapping peptides

CS877 (amino acids 257 to 271) and CS878 (amino acids 268 to 281), peptides CS880 (amino acids 286-299) and peptides CS882 (amino acids 315-333) were identified by hyperimmune sera.

5

EXAMPLE 9 - HSP70 (DnaK) from *Streptococcus pyogenes* and *Streptococcus agalactiae*: Molecular Cloning and DNA Sequencing of the *hsp70* Genes; Nucleotide and Protein Sequence Analyses; Antigenic Relatedness to *S. pneumoniae*;
10 Increased *Streptococcus agalactiae* HSP70 synthesis in response to heat.

A. Procedures

1. Bacterial Strains and Plasmid Vector

15 The strains of *S. pyogenes* (Group A *Streptococcus*) and *S. agalactiae* (Group B *Streptococcus*) used in this study were provided by the Laboratoire de la Santé Publique du Québec (LSPQ), Sainte-Anne de Bellevue, Québec, Canada. *S. agalactiae* type II strain V8
20 corresponds to the ATCC strain 12973. *S. pyogenes* strain Bruno corresponds to the ATCC strain 19615. The *E. coli* strain XLI Blue MRF' was obtained from Stratagene.

Streptococcal strains were grown at 37°C in a 5 % CO₂ incubator. The streptococci were streaked on
25 tryptic soy agar plates containing 5 % sheep blood (Les Laboratoires Quélab, Montréal, Canada), liquid cultures were made in heart infusion broth (Difco Laboratories, Detroit, MI) without agitation. The *E. coli* strain was grown at 37°C in L-broth with agitation at 250 rpm or on L-
30 agar.

The general cloning phagemid pBluescript KS(-) was purchased from Stratagene.

2. Recombinant DNA Techniques

35 Restriction enzymes, T4 DNA ligase, and calf intestinal phosphatase were used as recommended by the suppliers (Pharmacia [Canada] Inc., Baie d'Urfe, Canada; and New England Biolabs Ltd., Mississauga, Canada).

Preparation of plasmids by equilibrium centrifugation in - CsCl-ethidium bromide gradients, agarose gel electrophoresis of DNA fragments, Southern hybridization, and colony DNA hybridization were performed as described
5 by J. Sambrook et al. [supra]. Chromosomal DNA of the streptococcal bacteria was prepared using the procedure of B. M. Jayarao et al. [J. Clin. Microbiol., 29, pp. 2774-2778 (1991)] adapted for bacterial cultures of 90 ml. Rapid plasmid preparations were made accordingly to D.
10 Ish-Horowicz et al. [Nucl. Acids Res. 9, pp. 2989-2998 (1981)]. Plasmids used for DNA sequencing were purified using plasmid kits from Qiagen Inc. (Chatsworth, CA). DNA fragments were purified from agarose gels by the method of phenol freeze [S. A. Benson, Biotechniques 2, pp. 67-68
15 (1984)]. DNA probes were labeled with a³²P-dCTP or digoxigenin (DIG)-11-dUTP using the random primer labeling kits of Boehringer Mannheim (Laval, Canada). Plasmid transformations were carried out by the method of Simanis [Hanahan, D. In D. M. Glover (ed.), DNA Cloning, pp. 109-
20 135, (1985)]. The sequencing of genomic DNA inserts in plasmids was done using synthetic oligonucleotides. The sequencing reactions were carried out by the polymerase chain reaction (PCR) using the Taq Dye Deoxy Terminator Cycle Sequencing kit (ABI) and DNA electrophoresis was
25 performed on automated DNA sequencer 373A (ABI). The assembly of the DNA sequence was performed using the program Sequencher 3.0 from the Gene Codes Corporation (Ann Arbor, MI). Analysis of the DNA sequences and their predicted polypeptides were performed with the program
30 Gene Works version 2.45 from Intelligenetics, Inc. (Mountain View, CA). DNA amplification reactions were made using a DNA Thermal Cycler 480, Perkin Elmer. Oligonucleotides were synthesized by oligonucleotide synthesizer model 394 (ABI).

3. Molecular Cloning of the Genes *hsp70/dnak* of *S. agalactiae* and *S. pyogenes*

Chromosomal DNA from *S. agalactiae* and *S. pyogenes* was digested to completion with various restriction enzymes with palindromic hexanucleotide recognition sequences. The digests were analysed by Southern hybridization using a labeled PCR-amplified DNA probe corresponding to a 782 base-pairs region starting at base 332 downstream from the ATG initiation codon of the HSP72 gene of *S. pneumoniae* (see SEQ ID NO 4). This DNA region was selected because it is relatively well conserved among the *hsp70* genes of Gram-positive bacteria that have been characterized. The PCR amplification was done on the genomic DNA of *S. pneumoniae* using the oligonucleotides OCRR2 (5'-AAGCTGTTATCACAGTTCCGG) and OCRR3 (5'-GATACCAAGTGACAATGGCG). Hybridizing genomic restriction fragments of sufficient size to code for a 70-kDa polypeptide (>1.8 kb) were partially purified by extraction of genomic fragments of corresponding size from agarose gel. Verification of the presence of the *hsp70* gene among the purified genomic restriction fragments was done by Southern hybridization using the labeled 782-bp *S. pneumoniae* DNA probe.

The purified genomic DNA restriction fragments were cloned into dephosphorylated compatible restriction sites of pBluescript KS(-) and transformed into the *E. coli* strain XLI Blue MRF'. The colonies were screened by DNA hybridization using the labeled 782-bp *S. pneumoniae* DNA probe. Extracted plasmids were digested with various restriction enzymes to evaluate the size of the inserts and to verify the presence of the *hsp70* gene by Southern hybridization using the labeled 782-bp *S. pneumoniae* DNA probe. Plasmid pURV5 contains a 4.2-kb *HindIII* insert of the genomic DNA of *S. agalactiae*. Plasmid pURV4 contains a 3.5-kb *HindIII* fragment of the genomic DNA of *S. pyogenes*.

4. Heat Shock and Protein Labeling

The stress response of *S. agalactiae* to an heat shock was assayed by pulse-labeling with [³⁵S]methionine as described before in Example 1. *S. agalactiae* bacteria grown overnight in SMAM (Methionine assay Medium supplemented with 1 mg/l methionine, 1% (v/v) Isovitalex and 1 mg/l choline chloride) were pelleted by centrifugation and then resuspended in the methionine-free SMAM medium. The bacteria were incubated at 37°C for 1 h and then divided into two fractions of equal volume. The samples were either incubated at 37 or 43°C for 10 minutes and then labeled with 100 µCi/ml [³⁵S]methionine for 30 minutes at 37°C. The bacteria were extensively washed with PBS and cell extracts were prepared by treatment with mutanolysine and lysozyme as described for the DNA isolation (M.Jayarao et al., supra) followed by sonication.

5. Immunological Characterization

A series of six monoclonal antibodies raised to the HSP72_{rec} protein (F3-Pn3.5 to F3-Pn3.10) and the monoclonal antibodies F1-Pn3.1, F2-Pn3.2, F2-Pn3.3, F2-Pn3.4 were tested for their reactivity to HSP70 antigens from *S. pyogenes* and *S. agalactiae* by Western blot analysis. Cell lysates from *S. pyogenes* and *S. agalactiae* were obtained from treatment with mutanolysine and lysozyme (M.Jayarao et al., supra), sonication and boiling in SDS-PAGE sample buffer. Cell lysates from *E. coli* transformed with either pURV4 or pURV6 producing truncated *S. pyogenes* HSP70 antigens were tested after boiling in SDS-PAGE sample buffer.

B. DNA Sequence Analysis of the hsp70 /dnak Genes of Streptococcus pyogenes, Streptococcus agalactiae and Streptococcus pneumoniae

A region of 2438 bases in the 4.2-kb HindIII insert of plasmid pURV5 was sequenced. This sequence

contains an open reading frame (ORF) of 1830 nucleotides coding for a polypeptide of 609 amino acids with a molecular weight of 64907 (see SEQ ID NO: 7). The ORF has an ATG start codon beginning at position 248 and TAA stop codon ending at position 2077. The ATG start codon is preceded by the sequence GAGG, starting at position 237, which is complementary to 16S rRNA and serves as a ribosome binding site in *E. coli* [G. D. Stormo et al., Nucleic Acids Res. 10, pp. 2971-2996 (1982)]. The ORF and the polypeptide of the HSP70 of *S. agalactiae* are, respectively, identical at 85 and 95 % to the ORF and polypeptide of the HSP72 of *S. pneumoniae*.

Preliminary sequence comparisons with the HSP72 of *S. pneumoniae* showed that the 3.5-kb *HindIII* insert in plasmid pURV4 lacks the 3'-end coding region of the *hsp70* of *S. pyogenes*. An attempt to clone a 3-kb *SalI* genomic fragment containing the entire coding region of *hsp70* of *S. pyogenes* yielded plasmid pURV6 containing a 3.1-kb insert lacking the 5'-end coding region of the gene. The assembly of the *hsp70* gene regions present in plasmids pURV4 and pURV6 gave a 2183 nucleotide region containing an ORF of 1824 bases coding for a polypeptide of 608 amino acids with a molecular weight of 64847 (see SEQ ID NO: 20). The ATG start codon begins at position 204 and the TAA stop codon extends to position 2030. Similarly to the *hsp70* of *S. agalactiae*, the ATG start codon is preceded by a putative ribosome binding site sequence GAGG starting at position 193 [G. D. Stormo, supra]. The ORF and the deduced polypeptide of the *hsp70* of *S. pyogenes* are, respectively, identical at 85 and 94 % to the ORF and polypeptide of the HSP72 of *S. pneumoniae*. The ORF of plasmid pURV4 lacks 125 base pairs coding for 41 amino acids at the carboxyl end of the HSP70 of *S. pyogenes*; the ORF thus codes for the 567 amino acids of the amino end of that HSP70 (N-567_{rec}). The ORF of plasmid pURV6 lacks 114 base pairs coding for 38 amino acids at the amino end of the HSP70 of *S. pyogenes*; the ORF thus codes

for the 570 amino acids of the carboxyl end of that HSP70 (C-570_{rec}).

The global comparison of the DNA open reading frames (FIG. 24) and amino acid sequences (FIG. 25) of the HSP70/DnaK of *S. pyogenes*, *S. agalactiae*, and *S. pneumoniae* gave percentages of identity of 82 and 93 %, respectively.

C. Increased Synthesis of HSP70 by *S. agalactiae* in Response to Heat

One dimensional SDS-polyacrylamide gel electrophoretic analysis of cell extracts of heat-shocked and control *S. agalactiae* pulse-labeled with [³⁵S]methionine revealed that the synthesis of a 70 kDa-protein was significantly increased after a thermal stress (FIG. 26, lanes 1 and 2). Radioimmunoprecipitation analysis revealed that the heat inducible 70kDa-protein was easily detected at 43°C using monoclonal antibody F2-Pn3.4 thus indicating that the protein belongs to the heat shock protein 70 (hsp70/DnaK) family (FIG. 26, lanes 3 and 4).

D. Antigenic Relatedness of HSP70 Proteins in *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*

In this study, a panel of MAbs were used to investigate the antigenic relatedness of *S. pyogenes*, *S. agalactiae* and *S. pneumoniae* HSP70 proteins. Eight of ten MAbs reacted with all three *Streptococcus* species thus indicating that some B-cell epitopes are widely distributed among *S. pneumoniae*, *S. pyogenes* and *S. agalactiae*. The MAb F1-Pn3.1 which is directed against an epitope located between amino acid residues 584 and 607 of HSP72 from *S. pneumoniae* did not react with HSP70 antigens from either *S. pyogenes* or *S. agalactiae*. Comparison of this region among the three *Streptococcus* species revealed differences in 5 to 8 amino acids located between amino acids 589 and 596. The MAb F2-Pn3.3 which

was also directed against epitopes present in the C-151 region was reactive with *S. agalactiae* but not with *S. pyogenes*. These data clearly indicate that HSP70 proteins from *Streptococcus* species are structurally and immunologically related. There is however immunological distinction.

Analysis of the reactivity of MAbs F3-Pn3.5, F3-Pn3.6, F3-Pn3.7 and F3-Pn3.10 with truncated recombinant *S. pyogenes* HSP70 antigens allowed the identification of an antigenic region near the amino-terminal end on the *S. pneumoniae* HSP72. These MAbs reacted with constructs expressing the N-terminal 567 amino acid residues but failed to react with constructs expressing the C-570 fragment. These data localized the epitopes recognized by the MAbs F3-Pn3.5, F3-Pn3.6, F3-Pn3.7 and F3-Pn3.10 to between residues 1 and 38 of the HSP72 protein.

EXAMPLE 10 - Use of HSP70/HSP72 As A Human Vaccine

To formulate a vaccine for human use, appropriate HSP72 antigens may be selected from the polypeptides described herein. For example, one of skill in the art could design a vaccine around the HSP70/HSP72 polypeptide or fragments thereof containing an immunogenic epitope. The use of molecular biology techniques is particularly well-suited for the preparation of substantially pure recombinant antigens.

The vaccine composition may take a variety of forms. These include, for example solid, semi-solid and liquid dosage forms, such as powders, liquid solutions or suspensions, and liposomes. Based on our belief that the HSP70/HSP72 antigens of this invention may elicit a protective immune response when administered to a human, the compositions of this invention will be similar to those used for immunizing humans with other proteins and polypeptides, e.g. tetanus and diphtheria. Therefore, the

compositions of this invention will preferably comprise a pharmaceutically acceptable adjuvant such as incomplete Freund's adjuvant, aluminum hydroxide, a muramyl peptide, a water-in oil emulsion, a liposome, an ISCOM or CTB, or a
5 non-toxic B subunit from cholera toxin. Most preferably, the compositions will include a water-in-oil emulsion or aluminum hydroxide as adjuvant.

The composition would be administered to the patient in any of a number of pharmaceutically acceptable
10 forms including intramuscular, intradermal, subcutaneous or topic. Preferably, the vaccine will be administered intramuscularly.

Generally, the dosage will consist of an initial injection, most probably with adjuvant, of about 0.01 to
15 10 mg, and preferably 0.1 to 1.0 mg HSP72 antigen per patient, followed most probably by one or more booster injections. Preferably, boosters will be administered at about 1 and 6 months after the initial injection.

An important consideration relating to
20 pneumococcal vaccine development is the question of mucosal immunity. The ideal mucosal vaccine will be safely taken orally or intranasally as one or a few doses and would elicit protective antibodies on the appropriate surfaces along with systemic immunity. The mucosal
25 vaccine composition may include adjuvants, inert particulate carriers or recombinant live vectors.

The anti-HSP72 antibodies of this invention are useful for passive immunotherapy and immunoprophylaxis of humans infected with *S. pneumoniae*, *S. pyogenes*, *S.*
30 *agalactiae* or related bacteria. The dosage forms and regimens for such passive immunization would be similar to those of other passive immunotherapies.

An antibody according to this invention is exemplified by a hybridoma producing MAb F1-Pn3.1
35 deposited in the American Type Culture Collection in Rockville, Maryland, USA on July 21, 1995, and identified

as Murine Hybridoma Cell Line, F1-Pn3.1. This deposit was assigned accession number HB 11960.

While we have described herein a number of embodiments of this invention, it is apparent that our
5 basic embodiments may be altered to provide other embodiments that utilize the compositions and processes of this invention. Therefore, it will be appreciated that the scope of this invention includes all alternative
10 embodiments and variations that are defined in the foregoing specification and by the claims appended hereto; and the invention is not to be limited by the specific embodiments which have been presented herein by way of example.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Hamel, Josee
Brodeur, Bernard R
Martin, Denis
10 Rioux, Clement
- (ii) TITLE OF INVENTION: STREPTOCOCCAL HEAT SHOCK PROTEINS
MEMBERS OF THE HSP70 FAMILY
- 15 (iii) NUMBER OF SEQUENCES: 26
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Goudreau Gage Dubuc & Martineau Walker
(B) STREET: 800 Place Victoria, Suite 3400, Stock
20 Exchange Tower
(C) CITY: Montreal
(D) STATE: Quebec
(E) COUNTRY: CANADA
(F) ZIP: H4Z1E9
- 25 (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
30 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
35 (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/472,534
(B) FILING DATE: 07-JUN-1995
- 40 (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US (PROVIS) 60/001,805
(B) FILING DATE: 04-AUG-1995
- 45 (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Leclerc/Dubuc/Prince, Alain/Jean/Gaetan
(C) REFERENCE/DOCKET NUMBER: BIOVAC2-PCT
- (ix) TELECOMMUNICATION INFORMATION:
50 (A) TELEPHONE: (514) 397-7400
(B) TELEFAX: (514) 397-4382

(2) INFORMATION FOR SEQ ID NO:1:

- 55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3167 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
60 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- 65 (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Streptococcus pneumoniae
- 70 (ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 30..755

(ix) FEATURE:

5

(A) NAME/KEY: CDS

(B) LOCATION: 771..2912

(D) OTHER INFORMATION: /product= "FucI/HSP72 (C-169)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

10 GAACCTTCATT TTTAGAAAGG AGTGAGTTT ATG TCT CAA GAT GAA AAA TTA ATT 53
Met Ser Gln Asp Glu Lys Leu Ile
1 5

15 CGT GAA CAG ATT TGT GAT GTT TGT CAT AAG ATG TGG CAA CTT GGT TGG 101
Arg Glu Gln Ile Cys Asp Val Cys His Lys Met Trp Gln Leu Gly Trp
10 15 20

20 GTT GCT GCT AAC GAT GGG AAT GTA TCT GTT CGA TTA GAT GAG GAT ACC 149
Val Ala Ala Asn Asp Gly Asn Val Ser Val Arg Leu Asp Glu Asp Thr
25 30 35 40

25 ATT CTT GCA ACA CCT ACT GGT ATC AGC AAA AGT TTT ATT ACA CCA GAA 197
Ile Leu Ala Thr Pro Thr Gly Ile Ser Lys Ser Phe Ile Thr Pro Glu
45 50 55

30 AAG CTG GTG AAG TTA AAT CTT AAA GGA GAG ATT TTA GAA GCA GAA GGT 245
Lys Leu Val Lys Leu Asn Leu Lys Gly Glu Ile Leu Glu Ala Glu Gly
60 65 70

35 GAT TAC TGT CCT TCT AGT GAA ATT AAA ATG CAC ATT CGG TGC TAC GAA 293
Asp Tyr Cys Pro Ser Ser Glu Ile Lys Met His Ile Arg Cys Tyr Glu
75 80 85

40 GAA CGT GAG GAT GTT CGT TCA GTT GTT CAC GCG CAT CCA CCG ATT GCA 341
Glu Arg Glu Asp Val Arg Ser Val Val His Ala His Pro Pro Ile Ala
90 95 100

45 ACA GGA TTT GCT CTT GCA CAC ATT CCT TTA GAT ACT TAT TCA CTA ATT 389
Thr Gly Phe Ala Leu Ala His Ile Pro Leu Asp Thr Tyr Ser Leu Ile
105 110 115 120

50 GAG AGC GCG ATT GTG GTT GGG GCA ATT CCT ATT ACC CCA TTT GGA GTA 437
Glu Ser Ala Ile Val Val Gly Ala Ile Pro Ile Thr Pro Phe Gly Val
125 130 135

55 CCG TCT ACA ATG GAA GTG CCA GAA GCA ATT ACA CCT TAT CTG CCC GAT 485
Pro Ser Thr Met Glu Val Pro Glu Ala Ile Thr Pro Tyr Leu Pro Asp
140 145 150

60 CAT GAT GTC ATG CTA TTA GAA AAT CAT GGA GCT CTG ACT GTC GGA AGC 533
His Asp Val Met Leu Leu Glu Asn His Gly Ala Leu Thr Val Gly Ser
155 160 165

65 GAT GTC ATT ACA GCA TAC TAC CGT ATG GAA ACT TTA GAA TTA GTC GCA 581
Asp Val Ile Thr Ala Tyr Tyr Arg Met Glu Thr Leu Glu Leu Val Ala
170 175 180

70 AAG ACA ACC TTC CAC GGA AGA ATG TTA CTT TCT ACA AAG GGC ATT GAG 629
Lys Thr Thr Phe His Gly Arg Met Leu Leu Ser Thr Lys Gly Ile Glu
185 190 195 200

75 GAG CAA GAA ATT GCT CGT CCG ACT TTA GAA CGT CTA TTC TCA ATG CGA 677
Glu Gln Glu Ile Ala Arg Pro Thr Leu Glu Arg Leu Phe Ser Met Arg
205 210 215

80 GAA AAT TAT AAG GTT ACA GGT CGT CAC CCA GGC TAC CGT AAA TAT AAT 725
Glu Asn Tyr Lys Val Thr Gly Arg His Pro Gly Tyr Arg Lys Tyr Asn
220 225 230

	GGC GAT GGT AGT ATA AAA GAA ACA AAA AAA TAAGAGGAAA GTATT ATG ATC	776	-
	Gly Asp Gly Ser Ile Lys Glu Thr Lys Lys Met Ile		
	235 240 1		
5	CAA CAT CCA CGT ATT GGG ATT CGT CCG ACT ATT GAT GGT CGT CGT CAA	824	
	Gln His Pro Arg Ile Gly Ile Arg Pro Thr Ile Asp Gly Arg Arg Gln		
	5 10 15		
10	GGT GTA CGC GAA TCA CTT GAA GTA CAA ACA ATG AAC ATG GCT AAA AGT	872	
	Gly Val Arg Glu Ser Leu Glu Val Gln Thr Met Asn Met Ala Lys Ser		
	20 25 30		
15	GTG GCA GAT TTG ATT TCA AGC ACA TTG AAA TAT CCA GAT GGG GAA CCT	920	
	Val Ala Asp Leu Ile Ser Ser Thr Leu Lys Tyr Pro Asp Gly Glu Pro		
	35 40 45 50		
20	GTG GAA TGT GTG ATT TCT CCA TCT ACC ATT GGT CGT GTT CCA GAG GCT	968	
	Val Glu Cys Val Ile Ser Pro Ser Thr Ile Gly Arg Val Pro Glu Ala		
	55 60 65		
25	GCA GCT TCC CAT GAG TTG TTT AAA AAA TCA AAT GTT TGC GCA ACA ATT	1016	
	Ala Ala Ser His Glu Leu Phe Lys Lys Ser Asn Val Cys Ala Thr Ile		
	70 75 80		
30	ACA GTT ACA CCA TGC TGG TGT TAT GGT AGT GAA ACT ATG GAT ATG TCT	1064	
	Thr Val Thr Pro Cys Trp Cys Tyr Gly Ser Glu Thr Met Asp Met Ser		
	85 90 95		
35	CCA GAT ATT CCT CAT GCT ATT TGG GGA TTT AAT GGG ACA GAA CGC CCA	1112	
	Pro Asp Ile Pro His Ala Ile Trp Gly Phe Asn Gly Thr Glu Arg Pro		
	100 105 110		
40	GGA GCT GTC TAT CTT GCA GCT GTA CTA GCT TCA CAT ACT CAA AAA GGG	1160	
	Gly Ala Val Tyr Leu Ala Ala Val Leu Ala Ser His Thr Gln Lys Gly		
	115 120 125 130		
45	ATT CCA GCC TTT GGG ATT TAT GGT AGA GAT GTT CAG GAA GCT AAT GAT	1208	
	Ile Pro Ala Phe Gly Ile Tyr Gly Arg Asp Val Gln Glu Ala Asn Asp		
	135 140 145		
50	ACG GCT ATT CCA GAA GAT GTC AAA GAA AAA CTT TTA CGT TAT GCG CGG	1256	
	Thr Ala Ile Pro Glu Asp Val Lys Glu Lys Leu Leu Arg Tyr Ala Arg		
	150 155 160		
55	GCA GTT CTT GCA ACT GGC TTG ATG AGA GAC ACT GCT TAC CTA TCA ATG	1304	
	Ala Val Leu Ala Thr Gly Leu Met Arg Asp Thr Ala Tyr Leu Ser Met		
	165 170 175		
60	GGT AGT GTT TCG ATG GGG ATT GGT GGT TCT ATT GTA AAT CCA GAT TTC	1352	
	Gly Ser Val Ser Met Gly Ile Gly Gly Ser Ile Val Asn Pro Asp Phe		
	180 185 190		
65	TTC CAA GAA TAC TTA GGA ATG CGA AAT GAA TCG GTA GAT ATG ACG GAG	1400	
	Phe Gln Glu Tyr Leu Gly Met Arg Asn Glu Ser Val Asp Met Thr Glu		
	195 200 205 210		
70	TTC ACG CGC CGT ATG GAC CGT GGT ATT TAC GAC CCT GAA GAG TTC GAA	1448	
	Phe Thr Arg Arg Met Asp Arg Gly Ile Tyr Asp Pro Glu Glu Phe Glu		
	215 220 225		
75	CGT GCG CTC AAA TGG GTG AAA GAA AAC GTA AAA GAA GGA TTC GAC CAT	1496	
	Arg Ala Leu Lys Trp Val Lys Glu Asn Val Lys Glu Gly Phe Asp His		
	230 235 240		
80	AAC CGT GAA GAC CTT GTT TTA AGC CGT GAA GAA AAA GAT AGA CAA TGG	1544	
	Asn Arg Glu Asp Leu Val Leu Ser Arg Glu Glu Lys Asp Arg Gln Trp		
	245 250 255		

	GAA TTT GTT ATT AAG ATG TTC ATG ATT GGA CGT GAC TTA ATG GTT GGT	1592	-
	Glu Phe Val Ile Lys Met Phe Met Ile Gly Arg Asp Leu Met Val Gly		
	260 265 270		
5	AAC CCA AGA CTT GCT GAA CTT GGT TTT GAG GAA GAA GCA GTT GGT CAC	1640	
	Asn Pro Arg Leu Ala Glu Leu Gly Phe Glu Glu Glu Ala Val Gly His		
	275 280 285 290		
10	CAT GCT TTA GTA GCT GGT TTC CAA GGT CAA CGT CAG TGG ACA GAC CAT	1688	
	His Ala Leu Val Ala Gly Phe Gln Gly Gln Arg Gln Trp Thr Asp His		
	295 300 305		
15	TTT CCA AAT GGG GAC TTT ATG GAA ACT TTC CTC AAT ACT CAG TTT GAC	1736	
	Phe Pro Asn Gly Asp Phe Met Glu Thr Phe Leu Asn Thr Gln Phe Asp		
	310 315 320		
20	TGG AAT GGT ATT CGA AAA CCA TTT GTA TTT GCG ACA GAG AAT GAT TCA	1784	
	Trp Asn Gly Ile Arg Lys Pro Phe Val Phe Ala Thr Glu Asn Asp Ser		
	325 330 335		
25	CTA AAT GGT GTG TCT ATG CTC TTT AAT TAT CTA TTA ACA AAT ACT CCA	1832	
	Leu Asn Gly Val Ser Met Leu Phe Asn Tyr Leu Leu Thr Asn Thr Pro		
	340 345 350		
30	CAA ATC TTT GCT GAT GTG CGT ACT TAT TGG AGT CCA GAG GCT GTT GAA	1880	
	Gln Ile Phe Ala Asp Val Arg Thr Tyr Trp Ser Pro Glu Ala Val Glu		
	355 360 365 370		
35	CGT GTA ACA GGA TAT ACT TTA GAG GGT CGT GCT GCA GCT GGA TTC TTA	1928	
	Arg Val Thr Gly Tyr Thr Leu Glu Gly Arg Ala Ala Ala Gly Phe Leu		
	375 380 385		
40	CAT CTA ATC AAC TCT GGA TCT TGT ACA TTG GAT GGT ACA GGT CAA GCT	1976	
	His Leu Ile Asn Ser Gly Ser Cys Thr Leu Asp Gly Thr Gly Gln Ala		
	390 395 400		
45	ACT CGA GAT GGC AAA CCT GTT ATG AAA CCA TTC TGG GAG TTG GAT GAA	2024	
	Thr Arg Asp Gly Lys Pro Val Met Lys Pro Phe Trp Glu Leu Asp Glu		
	405 410 415		
50	AGT GAA GTA CAG GCT ATG CTT GAA AAT ACA GAC TTC CCA CCA GCA AAC	2072	
	Ser Glu Val Gln Ala Met Leu Glu Asn Thr Asp Phe Pro Pro Ala Asn		
	420 425 430		
55	CGC GAA TAC TTC CGT GGA GGA GGA TTC TCA ACT CGT TTC TTG ACG AAG	2120	
	Arg Glu Tyr Phe Arg Gly Gly Gly Phe Ser Thr Arg Phe Leu Thr Lys		
	435 440 445 450		
60	GGG GAT ATG CCA GTA ACA ATG GTA CGT CTC AAT CTT TTA AAA GGG GTT	2168	
	Gly Asp Met Pro Val Thr Met Val Arg Leu Asn Leu Leu Lys Gly Val		
	455 460 465		
65	GGT CCA GTG CTA CAA ATT GCA GAA GGT TAC ACA CTT GAA CTT CCT GAA	2216	
	Gly Pro Val Leu Gln Ile Ala Glu Gly Tyr Thr Leu Glu Leu Pro Glu		
	470 475 480		
70	GAT GTT CAC CAT ACT TTA GAT AAT CGT ACA GAT CCA GGA TGG CCA ACT	2264	
	Asp Val His His Thr Leu Asp Asn Arg Thr Asp Pro Gly Trp Pro Thr		
	485 490 495		
75	ACT TGG TTT GCT CCA CGT TTG ACA GGA AAA GGT GCT TTC AAG TCT GTC	2312	
	Thr Trp Phe Ala Pro Arg Leu Thr Gly Lys Gly Ala Phe Lys Ser Val		
	500 505 510		
80	TAT GAC GTC ATG AAT AAT TGG GGA GCT AAT CAC GGA GCC ATA ACA TAT	2360	
	Tyr Asp Val Met Asn Asn Trp Gly Ala Asn His Gly Ala Ile Thr Tyr		
	515 520 525 530		

	GGA CAC ATT GGA GCA GAC TTG ATT ACC TTG GCT TCT ATG TTG AGA ATT	2408	-
	Gly His Ile Gly Ala Asp Leu Ile Thr Leu Ala Ser Met Leu Arg Ile		
	535 540 545		
5	CCT CAA ATC GAA GTA ACA TTT GAC ATC GAC AAG AAC GGT ATC GTG TCT	2456	
	Pro Gln Ile Glu Val Thr Phe Asp Ile Asp Lys Asn Gly Ile Val Ser		
	550 555 560		
10	GTT AAG GCC AAA GAC CTT GGA ACT CAA AAA GAA CAA ACT ATT GTC ATC	2504	
	Val Lys Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile Val Ile		
	565 570 575		
15	CAA TCG AAC TCA GGT TTG ACT GAC GAA GAA ATC GAC CGC ATG ATG AAA	2552	
	Gln Ser Asn Ser Gly Leu Thr Asp Glu Glu Ile Asp Arg Met Met Lys		
	580 585 590		
20	GAT GCA GAA GCA AAC GCT GAA TCC GAT AAG AAA CGT AAA GAA GAA GTA	2600	
	Asp Ala Glu Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu Glu Val		
	595 600 605 610		
25	GAC CTT CGT AAT GAA GTG GAC CAA GCA ATC TTT GCG ACT GAA AAG ACA	2648	
	Asp Leu Arg Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr		
	615 620 625		
	ATC AAG GAA ACT GAA GGT AAA GGC TTC GAC GCA GAA CGT GAC GCT GCC	2696	
	Ile Lys Glu Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg Asp Ala Ala		
	630 635 640		
30	CAA GCT GCC CTT GAT GAC CTT AAG AAA GCT CAA GAA GAC AAC AAC TTG	2744	
	Gln Ala Ala Leu Asp Asp Leu Lys Lys Ala Gln Glu Asp Asn Asn Leu		
	645 650 655		
35	GAC GAC ATG AAA GCA AAA CTT GAA GCA TTG AAC GAA AAA GCT CAA GGA	2792	
	Asp Asp Met Lys Ala Lys Leu Glu Ala Leu Asn Glu Lys Ala Gln Gly		
	660 665 670		
40	CTT GCT GTT AAA CTC TAC GAA CAA GCC GCA GCA GCG CAA CAA GCT CAA	2840	
	Leu Ala Val Lys Leu Tyr Glu Gln Ala Ala Ala Gln Gln Ala Gln		
	675 680 685 690		
	GAA GGA GCA GAA GGC GCA CAA GCA ACA GGA AAC GCA GGC GAT GAC GTC	2888	
	Glu Gly Ala Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp Asp Val		
	695 700 705		
45	GTA GAC GGA GAG TTT ACG GAA AAG TAAGATGAGT GTATTGGATG AAGAGTATCT	2942	
	Val Asp Gly Glu Phe Thr Glu Lys		
	710		
50	AAAAAATACA CGAAAAGTTT ATAATGATTT TTGTAATCAA GCTGATAACT ATAGAACATC	3002	
	AAAAGATTTT ATTGATAATA TTCCAATAGA ATATTTAGCT AGATATAGAG AAATTATATT	3062	
55	AGCTGAGCAT GATAGTTGTG TCAAAAATGA TGAAGCGGTA AGGAATTTTG TTACCTCAGT	3122	
	ATTGTTGTCT GCATTTGTAT CGGCGATGGT ATCAGCTATG ATATC	3167	

(2) INFORMATION FOR SEQ ID NO:2:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 242 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

65

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

70

Met Ser Gln Asp Glu Lys Leu Ile Arg Glu Gln Ile Cys Asp Val Cys
 1 5 10 15

His Lys Met Trp Gln Leu Gly Trp Val Ala Ala Asn Asp Gly Asn Val
 20 25 30
 5 Ser Val Arg Leu Asp Glu Asp Thr Ile Leu Ala Thr Pro Thr Gly Ile
 35 40 45
 Ser Lys Ser Phe Ile Thr Pro Glu Lys Leu Val Lys Leu Asn Leu Lys
 50 55 60
 10 Gly Glu Ile Leu Glu Ala Glu Gly Asp Tyr Cys Pro Ser Ser Glu Ile
 65 70 75 80
 Lys Met His Ile Arg Cys Tyr Glu Glu Arg Glu Asp Val Arg Ser Val
 85 90 95
 Val His Ala His Pro Pro Ile Ala Thr Gly Phe Ala Leu Ala His Ile
 100 105 110
 20 Pro Leu Asp Thr Tyr Ser Leu Ile Glu Ser Ala Ile Val Val Gly Ala
 115 120 125
 Ile Pro Ile Thr Pro Phe Gly Val Pro Ser Thr Met Glu Val Pro Glu
 130 135 140
 25 Ala Ile Thr Pro Tyr Leu Pro Asp His Asp Val Met Leu Leu Glu Asn
 145 150 155 160
 His Gly Ala Leu Thr Val Gly Ser Asp Val Ile Thr Ala Tyr Tyr Arg
 165 170 175
 Met Glu Thr Leu Glu Leu Val Ala Lys Thr Thr Phe His Gly Arg Met
 180 185 190
 35 Leu Leu Ser Thr Lys Gly Ile Glu Glu Gln Glu Ile Ala Arg Pro Thr
 195 200 205
 Leu Glu Arg Leu Phe Ser Met Arg Glu Asn Tyr Lys Val Thr Gly Arg
 210 215 220
 40 His Pro Gly Tyr Arg Lys Tyr Asn Gly Asp Gly Ser Ile Lys Glu Thr
 225 230 235 240
 Lys Lys
 45

(2) INFORMATION FOR SEQ ID NO:3:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 714 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

60 Met Ile Gln His Pro Arg Ile Gly Ile Arg Pro Thr Ile Asp Gly Arg
 1 5 10 15
 Arg Gln Gly Val Arg Glu Ser Leu Glu Val Gln Thr Met Asn Met Ala
 20 25 30
 65 Lys Ser Val Ala Asp Leu Ile Ser Ser Thr Leu Lys Tyr Pro Asp Gly
 35 40 45
 Glu Pro Val Glu Cys Val Ile Ser Pro Ser Thr Ile Gly Arg Val Pro
 50 55 60

Glu Ala Ala Ala Ser His Glu Leu Phe Lys Lys Ser Asn Val Cys Ala
 65 70 75 80
 5 Thr Ile Thr Val Thr Pro Cys Trp Cys Tyr Gly Ser Glu Thr Met Asp
 85 90 95
 Met Ser Pro Asp Ile Pro His Ala Ile Trp Gly Phe Asn Gly Thr Glu
 100 105 110
 10 Arg Pro Gly Ala Val Tyr Leu Ala Ala Val Leu Ala Ser His Thr Gln
 115 120 125
 Lys Gly Ile Pro Ala Phe Gly Ile Tyr Gly Arg Asp Val Gln Glu Ala
 130 135 140
 15 Asn Asp Thr Ala Ile Pro Glu Asp Val Lys Glu Lys Leu Leu Arg Tyr
 145 150 155 160
 20 Ala Arg Ala Val Leu Ala Thr Gly Leu Met Arg Asp Thr Ala Tyr Leu
 165 170 175
 Ser Met Gly Ser Val Ser Met Gly Ile Gly Gly Ser Ile Val Asn Pro
 180 185 190
 25 Asp Phe Phe Gln Glu Tyr Leu Gly Met Arg Asn Glu Ser Val Asp Met
 195 200 205
 Thr Glu Phe Thr Arg Arg Met Asp Arg Gly Ile Tyr Asp Pro Glu Glu
 210 215 220
 Phe Glu Arg Ala Leu Lys Trp Val Lys Glu Asn Val Lys Glu Gly Phe
 225 230 235 240
 35 Asp His Asn Arg Glu Asp Leu Val Leu Ser Arg Glu Glu Lys Asp Arg
 245 250 255
 Gln Trp Glu Phe Val Ile Lys Met Phe Met Ile Gly Arg Asp Leu Met
 260 265 270
 40 Val Gly Asn Pro Arg Leu Ala Glu Leu Gly Phe Glu Glu Glu Ala Val
 275 280 285
 Gly His His Ala Leu Val Ala Gly Phe Gln Gly Gln Arg Gln Trp Thr
 290 295 300
 Asp His Phe Pro Asn Gly Asp Phe Met Glu Thr Phe Leu Asn Thr Gln
 305 310 315 320
 50 Phe Asp Trp Asn Gly Ile Arg Lys Pro Phe Val Phe Ala Thr Glu Asn
 325 330 335
 Asp Ser Leu Asn Gly Val Ser Met Leu Phe Asn Tyr Leu Leu Thr Asn
 340 345 350
 55 Thr Pro Gln Ile Phe Ala Asp Val Arg Thr Tyr Trp Ser Pro Glu Ala
 355 360 365
 Val Glu Arg Val Thr Gly Tyr Thr Leu Glu Gly Arg Ala Ala Ala Gly
 370 375 380
 Phe Leu His Leu Ile Asn Ser Gly Ser Cys Thr Leu Asp Gly Thr Gly
 385 390 395 400
 65 Gln Ala Thr Arg Asp Gly Lys Pro Val Met Lys Pro Phe Trp Glu Leu
 405 410 415
 Asp Glu Ser Glu Val Gln Ala Met Leu Glu Asn Thr Asp Phe Pro Pro
 420 425 430

Ala Asn Arg Glu Tyr Phe Arg Gly Gly Gly Phe Ser Thr Arg Phe Leu
435 440 445

5 Thr Lys Gly Asp Met Pro Val Thr Met Val Arg Leu Asn Leu Leu Lys
450 455 460

Gly Val Gly Pro Val Leu Gln Ile Ala Glu Gly Tyr Thr Leu Glu Leu
465 470 475 480

10 Pro Glu Asp Val His His Thr Leu Asp Asn Arg Thr Asp Pro Gly Trp
485 490 495

Pro Thr Thr Trp Phe Ala Pro Arg Leu Thr Gly Lys Gly Ala Phe Lys
500 505 510

15 Ser Val Tyr Asp Val Met Asn Asn Trp Gly Ala Asn His Gly Ala Ile
515 520 525

Thr Tyr Gly His Ile Gly Ala Asp Leu Ile Thr Leu Ala Ser Met Leu
530 535 540

Arg Ile Pro Gln Ile Glu Val Thr Phe Asp Ile Asp Lys Asn Gly Ile
545 550 555 560

25 Val Ser Val Lys Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile
565 570 575

Val Ile Gln Ser Asn Ser Gly Leu Thr Asp Glu Glu Ile Asp Arg Met
580 585 590

30 Met Lys Asp Ala Glu Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu
595 600 605

Glu Val Asp Leu Arg Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu
610 615 620

Lys Thr Ile Lys Glu Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg Asp
625 630 635 640

40 Ala Ala Gln Ala Ala Leu Asp Asp Leu Lys Lys Ala Gln Glu Asp Asn
645 650 655

Asn Leu Asp Asp Met Lys Ala Lys Leu Glu Ala Leu Asn Glu Lys Ala
660 665 670

Gln Gly Leu Ala Val Lys Leu Tyr Glu Gln Ala Ala Ala Gln Gln
675 680 685

50 Ala Gln Glu Gly Ala Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp
690 695 700

Asp Val Val Asp Gly Glu Phe Thr Glu Lys
705 710

55

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4320 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
(A) ORGANISM: *Streptococcus pneumoniae*

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 682..2502
 (D) OTHER INFORMATION: /product= "Heat-shock protein 72"

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 3265..4320
 (D) OTHER INFORMATION: /product= "NH2-terminal portion of DNA J"

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 682..2502

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

20	AAGCTTGATT CACGCTTTGA AAGAAGAAGG AATTGAAGAA ATCGCAGCAG ATGGCGAATT	60
	TGACCATAAC TACCATATGG CCATCCAAAC TCTCCAGCA GACGATGAAC ACCCAGTAGA	120
25	TACCATCGCC CAAGTCTTTC AAAAAGGCTA CAACTCCAT GACCGCATCC TACGCCCAGC	180
	AATGGTAGTG GTGTATAACT AAGATACAAA GCCCGTAAAA AGCTCGCAGT AAAAATAGGA	240
	GATTGACGAA GTGTTGATG AACACAAGAA AATCTATCTT TTTTACTCAG AGCTTAGGGC	300
30	GTGTTGATG CGGCAATTCT GACGGTAGCT AAAGCAACTC GTCAGAAAAC GGCAGTCGCT	360
	ATGGCGTTTG TCTAGCTTCC TTTACTAACTC GTCGTCGAAA TAAAATCGAT TTCGACTCTT	420
35	CGTGTCGCAA TTTACATAAT AGAAAAGTTG TCCGAAACGA CAATAAACTA TGAAGAAAGA	480
	TAAAATATGT TTGGCTTTGT AATAGTGAGC GAAGCGAACC AAAGACGATA CTCTTCGCTG	540
	TGGCGCTATT TGCGCAAATT TTGAGACCTT AGGCTCAAAG TTTAGTCAAA GAGATTGACA	600
40	AAGTCAAGCT CTGACGGCGT CGCCACTTAA GAAGAGTATC AAAAAGAAAA ATAGAAAATT	660
	AACTAACAAG GAGAAAAACA C ATG TCT AAA ATT ATC GGT ATT GAC TTA GGT	711
	Met Ser Lys Ile Ile Gly Ile Asp Leu Gly	
	1 5 10	
45	ACA ACA AAC TCA GCA GTT GCA GTT CTT GAA GGA ACT GAA AGC AAA ATC	759
	Thr Thr Asn Ser Ala Val Ala Val Leu Glu Gly Thr Glu Ser Lys Ile	
	15 20 25	
50	ATC GCA AAC CCA GAA GGA AAC CGC ACA ACT CCA TCT GTA GTC TCA TTC	807
	Ile Ala Asn Pro Glu Gly Asn Arg Thr Thr Pro Ser Val Val Ser Phe	
	30 35 40	
55	AAA AAC GGA GAA ATC ATC GTT GGT GAT GCT GCA AAA CGT CAA GCA GTT	855
	Lys Asn Gly Glu Ile Ile Val Gly Asp Ala Ala Lys Arg Gln Ala Val	
	45 50 55	
60	ACA AAC CCA GAT ACA GTT ATC TCT ATC AAA TCT AAG ATG GGA ACT TCT	903
	Thr Asn Pro Asp Thr Val Ile Ser Ile Lys Ser Lys Met Gly Thr Ser	
	60 65 70	
65	GAA AAA GTT TCT GCA AAT GGA AAA GAA TAC ACT CCA CAA GAA ATC TCA	951
	Glu Lys Val Ser Ala Asn Gly Lys Glu Tyr Thr Pro Gln Glu Ile Ser	
	75 80 85 90	
	GCT ATG ATC CTT CAA TAC TTG AAA GGC TAC GCT GAA GAC TAC CTT GGT	999
	Ala Met Ile Leu Gln Tyr Leu Lys Gly Tyr Ala Glu Asp Tyr Leu Gly	
	95 100 105	

	GAG	AAA	GTA	ACC	AAA	GCT	GTT	ATC	ACA	GTT	CCG	GCT	TAC	TTC	AAC	GAC	1047	-
	Glu	Lys	Val	Thr	Lys	Ala	Val	Ile	Thr	Val	Pro	Ala	Tyr	Phe	Asn	Asp		
				110					115					120				
5	GCT	CAA	CGT	CAA	GCA	ACA	AAA	GAC	GCT	GGT	AAA	ATT	GCT	GGT	CTT	GAA	1095	
	Ala	Gln	Arg	Gln	Ala	Thr	Lys	Asp	Ala	Gly	Lys	Ile	Ala	Gly	Leu	Glu		
			125					130					135					
10	GTA	GAA	CGT	ATT	GTT	AAC	GAA	CCA	ACT	GCA	GCA	GCT	CTT	GCT	TAT	GGT	1143	
	Val	Glu	Arg	Ile	Val	Asn	Glu	Pro	Thr	Ala	Ala	Ala	Leu	Ala	Tyr	Gly		
			140				145					150						
15	TTG	GAC	AAG	ACT	GAC	AAA	GAA	GAA	AAA	ATC	TTG	GTA	TTT	GAC	CTT	GGT	1191	
	Leu	Asp	Lys	Thr	Asp	Lys	Glu	Glu	Lys	Ile	Leu	Val	Phe	Asp	Leu	Gly		
	155					160					165					170		
20	GGT	GGT	ACA	TTC	GAC	GTC	TCT	ATC	CTT	GAA	TTG	GGT	GAC	GGT	GTC	TTC	1239	
	Gly	Gly	Thr	Phe	Asp	Val	Ser	Ile	Leu	Glu	Leu	Gly	Asp	Gly	Val	Phe		
				175						180					185			
25	GAC	GTA	TTG	TCA	ACT	GCA	GGG	GAC	AAC	AAA	CTT	GGT	GGT	GAC	GAC	TTT	1287	
	Asp	Val	Leu	Ser	Thr	Ala	Gly	Asp	Asn	Lys	Leu	Gly	Gly	Asp	Asp	Phe		
				190					195					200				
30	GAC	CAA	AAA	ATC	ATT	GAC	CAC	TTG	GTA	GCA	GAA	TTC	AAG	AAA	GAA	AAC	1335	
	Asp	Gln	Lys	Ile	Ile	Asp	His	Leu	Val	Ala	Glu	Phe	Lys	Lys	Glu	Asn		
			205				210					215						
35	GGT	ATC	GAC	TTG	TCT	ACT	GAC	AAG	ATG	GCA	ATG	CAA	CGT	TTG	AAA	GAT	1383	
	Gly	Ile	Asp	Leu	Ser	Thr	Asp	Lys	Met	Ala	Met	Gln	Arg	Leu	Lys	Asp		
			220				225					230						
40	GCG	GCT	GAA	AAA	GCG	AAG	AAA	GAC	CTT	TCT	GGT	GTA	ACT	TCA	ACA	CAA	1431	
	Ala	Ala	Glu	Lys	Ala	Lys	Lys	Asp	Leu	Ser	Gly	Val	Thr	Ser	Thr	Gln		
	235				240					245						250		
45	ATC	AGC	TTG	CCA	TTT	ATC	ACT	GCA	GGT	GAG	GCT	GGA	CCT	CTT	CAC	TTG	1479	
	Ile	Ser	Leu	Pro	Phe	Ile	Thr	Ala	Gly	Glu	Ala	Gly	Pro	Leu	His	Leu		
				255						260				265				
50	GAA	ATG	ACT	TTA	ACT	CGT	GCG	AAA	TTT	GAT	GAT	TTG	ACT	CGT	GAC	CTT	1527	
	Glu	Met	Thr	Leu	Thr	Arg	Ala	Lys	Phe	Asp	Asp	Leu	Thr	Arg	Asp	Leu		
				270					275					280				
55	GTT	GAA	CGT	ACA	AAA	GTT	CCA	GTT	CGT	CAA	GCC	CTT	TCA	GAT	GCA	GGT	1575	
	Val	Glu	Arg	Thr	Lys	Val	Pro	Val	Arg	Gln	Ala	Leu	Ser	Asp	Ala	Gly		
			285				290					295						
60	TTG	AGC	TTG	TCA	GAA	ATC	GAC	GAA	GTT	ATC	CTT	GTT	GGT	GGT	TCA	ACT	1623	
	Leu	Ser	Leu	Ser	Glu	Ile	Asp	Glu	Val	Ile	Leu	Val	Gly	Gly	Ser	Thr		
			300				305					310						
65	CGT	ATC	CCT	GCC	GTT	GTT	GAA	GCT	GTT	AAA	GCT	GAA	ACT	GGT	AAA	GAA	1671	
	Arg	Ile	Pro	Ala	Val	Val	Glu	Ala	Val	Lys	Ala	Glu	Thr	Gly	Lys	Glu		
	315				320					325						330		
70	CCA	AAC	AAA	TCA	GTA	AAC	CCT	GAT	GAA	GTA	GTT	GCT	ATG	GGT	GCG	GCT	1719	
	Pro	Asn	Lys	Ser	Val	Asn	Pro	Asp	Glu	Val	Val	Ala	Met	Gly	Ala	Ala		
				335					340					345				
75	ATC	CAA	GGT	GGT	GTG	ATT	ACT	GGT	GAT	GTC	AAG	GAT	GTT	GTC	CTT	CTT	1767	
	Ile	Gln	Gly	Gly	Val	Ile	Thr	Gly	Asp	Val	Lys	Asp	Val	Val	Leu	Leu		
				350				355						360				
80	GAT	GTA	ACG	CCA	TTG	TCA	CTT	GGT	ATC	GAA	ACA	ATG	GGT	GGA	GTA	TTT	1815	
	Asp	Val	Thr	Pro	Leu	Ser	Leu	Gly	Ile	Glu	Thr	Met	Gly	Gly	Val	Phe		
			365				370						375					

5	ACA AAA CTT ATC GAT CGC AAC ACT ACA ATC CCA ACA TCT AAA TCA CAA Thr Lys Leu Ile Asp Arg Asn Thr Thr Ile Pro Thr Ser Lys Ser Gln 380 385 390	1863	-
10	GTC TTC TCA ACA GCA GCA GAC AAC CAA CCA GCC GTT GAT ATC CAC GTT Val Phe Ser Thr Ala Ala Asp Asn Gln Pro Ala Val Asp Ile His Val 395 400 405 410	1911	
15	CTT CAA GGT GAA CGC CCA ATG GCA GCA GAT AAC AAG ACT CTT GGA CGC Leu Gln Gly Glu Arg Pro Met Ala Ala Asp Asn Lys Thr Leu Gly Arg 415 420 425	1959	
20	TTC CAA TTG ACT GAT ATC CCA GCT GCA CCT CGT GGA ATT CCT CAA ATC Phe Gln Leu Thr Asp Ile Pro Ala Ala Pro Arg Gly Ile Pro Gln Ile 430 435 440	2007	
25	GAA GTA ACA TTT GAC ATC GAC AAG AAC GGT ATC GTG TCT GTT AAG GCC Glu Val Thr Phe Asp Ile Asp Lys Asn Gly Ile Val Ser Val Lys Ala 445 450 455	2055	
30	AAA GAC CTT GGA ACT CAA AAA GAA CAA ACT ATT GTC ATC CAA TCG AAC Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile Val Ile Gln Ser Asn 460 465 470	2103	
35	TCA GGT TTG ACT GAC GAA GAA ATC GAC CGC ATG ATG AAA GAT GCA GAA Ser Gly Leu Thr Asp Glu Glu Ile Asp Arg Met Met Lys Asp Ala Glu 475 480 485 490	2151	
40	GCA AAC GCT GAA TCC GAT AAG AAA CGT AAA GAA GAA GTA GAC CTT CGT Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu Glu Val Asp Leu Arg 495 500 505	2199	
45	AAT GAA GTG GAC CAA GCA ATC TTT GCG ACT GAA AAG ACA ATC AAG GAA Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr Ile Lys Glu 510 515 520	2247	
50	ACT GAA GGT AAA GGC TTC GAC GCA GAA CGT GAC GCT GCC CAA GCT GCC Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg Asp Ala Ala Gln Ala Ala 525 530 535	2295	
55	CTT GAT GAC CTT AAG AAA GCT CAA GAA GAC AAC AAC TTG GAC GAC ATG Leu Asp Asp Leu Lys Lys Ala Gln Glu Asp Asn Asn Leu Asp Asp Met 540 545 550	2343	
60	AAA GCA AAA CTT GAA GCA TTG AAC GAA AAA GCT CAA GGA CTT GCT GTT Lys Ala Lys Leu Glu Glu Ala Leu Asn Glu Lys Ala Gln Gly Leu Ala Val 555 560 565 570	2391	
65	AAA CTC TAC GAA CAA GCC GCA GCA GCG CAA CAA GCT CAA GAA GGA GCA Lys Leu Tyr Glu Gln Ala Ala Ala Ala Gln Gln Ala Gln Glu Gly Ala 575 580 585	2439	
70	GAA GGC GCA CAA GCA ACA GGA AAC GCA GGC GAT GAC GTC GTA GAC GGA Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp Asp Val Val Asp Gly 590 595 600	2487	
	GAG TTT ACG GAA AAG TAAGATGAGT GTATTGGATG AAGAGTATCT AAAAAATACA Glu Phe Thr Glu Lys 605	2542	
	CGAAAAGTTT ATAATGATTT TTGTAATCAA GCTGATAACT ATAGAACATC AAAAGATTTT	2602	
	ATTGATAATA TTCCAATAGA ATATTTAGCT AGATATAGAG AAATTATATT AGCTGAGCAT	2662	
	GATAGTTGTG TCAAAAATGA TGAAGCGGTA AGGAATTTTG TTACCTCAGT ATTGTTGTCT	2722	
	GCAATTTGTAT CGGCGATGGT ATCAGCTATG ATATCATTAG AAATACAAAC ATATAAATTT	2782	
	GTAATACCGT TCATAATTGG TATGATTTGG ACAGTAGTTG TATTTCTTAT GATCAATTGG	2842	

	AATTATATAG	GCAAATACTA	AGAAGAGACA	AAAATATATA	AATATTTCTG	TACTTATAGG	2902
	ATATTTAAAA	TCCAAATAAA	GTTAATTTAC	TTATTTGCAG	AGGTTGCAAC	CCAGCCTCTG	2962
5	TTTTTCGATA	AAAAGGGACG	GAATCTCATT	TGTTTGGGTT	TTGTCTCATC	AATAGAAAGG	3022
	AACAAAGAGT	GTTTCGTAAC	GAACACGGGT	TTCAGAAATTT	CTTACTAAAT	ATAAAAGAAA	3082
10	GGAATTGAAC	CCGACCTAAA	TGGTGGTTCG	ATTGAGAACA	TCAATAGAAA	GGAATAAGGG	3142
	TGTTTCGTAAC	TGAACACGGG	CTACGGACTG	TGCCAAAAAG	ATAGTTTTTTT	CTAGGACGTA	3202
	AGCGTCCGTC	GTCAAACTC	CTAGATGGCT	GTGTCCGTTT	GACGCCCTTT	GTATCTTGAA	3262
15	TT ATG AAC AAT ACT GAA TTT TAT GAT CGT CTG GGG GTA TCC AAA AAC						3309
	Met Asn Asn Thr Glu Phe Tyr Asp Arg Leu Gly Val Ser Lys Asn	1	5	10	15		
20	GCT TCG GCA GAC GAA ATC AAA AAG GCT TAT CGT AAG CTT TCC AAA AAA						3357
	Ala Ser Ala Asp Glu Ile Lys Lys Ala Tyr Arg Lys Leu Ser Lys Lys	20	25	30			
25	TAT CAC CCA GAT ATC AAC AAG GAG CCT GGT GCT GAG GAC AAG TAC AAG						3405
	Tyr His Pro Asp Ile Asn Lys Glu Pro Gly Ala Glu Asp Lys Tyr Lys	35	40	45			
30	GAA GTT CAA GAA GCC TAT GAG ACT TTG AGT GAC GAC CAA AAA CGT GCT						3453
	Glu Val Gln Glu Ala Tyr Glu Thr Leu Ser Asp Asp Gln Lys Arg Ala	50	55	60			
	GCC TAT GAC CAG TAT GGT GCT GCA GGC GCC AAT GGT GGT TTT GGT GGA						3501
	Ala Tyr Asp Gln Tyr Gly Ala Ala Gly Ala Asn Gly Gly Phe Gly Gly	65	70	75			
35	GCT GGT GGT TTC GGC GGT TTC AAT GGG GCA GGT GGC TTC GGT GGT TTT						3549
	Ala Gly Gly Phe Gly Gly Phe Asn Gly Ala Gly Gly Phe Gly Gly Phe	80	85	90	95		
40	GAG GAT ATT TTC TCA AGT TTC TTC GGC GGA GGC GGT TCT TCG CGC AAT						3597
	Glu Asp Ile Phe Ser Ser Phe Phe Gly Gly Gly Ser Ser Arg Asn	100	105	110			
45	CCA AAC GCT CCT CGC CAA GGA GAT GAT CTC CAG TAT CGT GTC AAT TTG						3645
	Pro Asn Ala Pro Arg Gln Gly Asp Leu Gln Tyr Arg Val Asn Leu	115	120	125			
50	ACC TTT GAA GAA GCT ATC TTC GGA ACT GAG AAG GAA GTT AAG TAT CAT						3693
	Thr Phe Glu Glu Ala Ile Phe Gly Thr Glu Lys Glu Val Lys Tyr His	130	135	140			
	CGT GAA GCT GGC TGT CGT ACA TGT AAT GGA TCT GGT GCT AAG CCA GGG						3741
	Arg Glu Ala Gly Cys Arg Thr Cys Asn Gly Ser Gly Ala Lys Pro Gly	145	150	155			
55	ACA AGT CCA GTC ACT TGT GGA CGC TGT CAT GGC GCT GGT GTC ATT AAC						3789
	Thr Ser Pro Val Thr Cys Gly Arg Cys His Gly Ala Gly Val Ile Asn	160	165	170	175		
60	GTC GAT ACG CAG ACT CCT CTT GGT ATG ATG CGT CGC CAA GTA ACC TGT						3837
	Val Asp Thr Gln Thr Pro Leu Gly Met Met Arg Arg Gln Val Thr Cys	180	185	190			
65	GAT GTC TGT CAC GGT CGA GGA AAA GAA ATC AAA TAT CCA TGT ACA ACC						3885
	Asp Val Cys His Gly Arg Gly Lys Glu Ile Lys Tyr Pro Cys Thr Thr	195	200	205			
	TGT CAT GGA ACA GGT CAT GAG AAA CAA GCT CAT AGC GTA CAT GTG AAA						3933
	Cys His Gly Thr Gly His Glu Lys Gln Ala His Ser Val His Val Lys	210	215	220			

	ATC CCT GCT GGT GTG GAA ACA GGT CAA CAA ATT CGC CTC GCT GGT CAA Ile Pro Ala Gly Val Glu Thr Gly Gln Gln Ile Arg Leu Ala Gly Gln 225 230 235	3981	-
5	GGT GAA GCA GGC TTT AAC GGT GGA CCT TAT GGT GAC TTG TAT GTA GTA Gly Glu Ala Gly Phe Asn Gly Gly Pro Tyr Gly Asp Leu Tyr Val Val 240 245 250 255	4029	
10	GTT TCT GTG GAA GCT AGT GAC AAG TTT GAA CGT GAA GGA ACG ACT ATC Val Ser Val Glu Ala Ser Asp Lys Phe Glu Arg Glu Gly Thr Thr Ile 260 265 270	4077	
15	TTC TAC AAT CTC AAC CTC AAC TTT GTC CAA GCG GCT CTT GGT GAT ACA Phe Tyr Asn Leu Asn Leu Asn Phe Val Gln Ala Ala Leu Gly Asp Thr 275 280 285	4125	
20	GTA GAT ATT CCA ACT GTT CAC GGT GAT GTT GAA TTG GTT ATT CCA GAG Val Asp Ile Pro Thr Val His Gly Asp Val Glu Leu Val Ile Pro Glu 290 295 300	4173	
25	GGA ACT CAG ACT GGT AAG AAA TTC CGC CTA CGT AGT AAG GGG GCA CCG Gly Thr Gln Thr Gly Lys Lys Phe Arg Leu Arg Ser Lys Gly Ala Pro 305 310 315	4221	
	AGC CTT CGT GGC GGT GCA GTT GGT GAC CAA TAC GTT ACT GTT AAT GTC Ser Leu Arg Gly Gly Ala Val Gly Asp Gln Tyr Val Thr Val Asn Val 320 325 330 335	4269	
30	GTA ACA CCG ACA GGC TTG AAC GAC CGC CAA AAA GTA GCC TTG AAA GAA Val Thr Pro Thr Gly Leu Asn Asp Arg Gln Lys Val Ala Leu Lys Glu 340 345 350	4317	
35	TTC Phe	4320	

40 (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 607 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

45 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

50	Met Ser Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Ala Val 1 5 10 15
	Ala Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro Glu Gly 20 25 30
55	Asn Arg Thr Thr Pro Ser Val Val Ser Phe Lys Asn Gly Glu Ile Ile 35 40 45
60	Val Gly Asp Ala Ala Lys Arg Gln Ala Val Thr Asn Pro Asp Thr Val 50 55 60
	Ile Ser Ile Lys Ser Lys Met Gly Thr Ser Glu Lys Val Ser Ala Asn 65 70 75 80
65	Gly Lys Glu Tyr Thr Pro Gln Glu Ile Ser Ala Met Ile Leu Gln Tyr 85 90 95
	Leu Lys Gly Tyr Ala Glu Asp Tyr Leu Gly Glu Lys Val Thr Lys Ala 100 105 110

Val Ile Thr Val Pro Ala Tyr Phe Asn Asp Ala Gln Arg Gln Ala Thr
 115 120 125
 5 Lys Asp Ala Gly Lys Ile Ala Gly Leu Glu Val Glu Arg Ile Val Asn
 130 135 140
 Glu Pro Thr Ala Ala Ala Leu Ala Tyr Gly Leu Asp Lys Thr Asp Lys
 145 150 155 160
 10 Glu Glu Lys Ile Leu Val Phe Asp Leu Gly Gly Gly Thr Phe Asp Val
 165 170 175
 Ser Ile Leu Glu Leu Gly Asp Gly Val Phe Asp Val Leu Ser Thr Ala
 180 185 190
 15 Gly Asp Asn Lys Leu Gly Gly Asp Asp Phe Asp Gln Lys Ile Ile Asp
 195 200 205
 20 His Leu Val Ala Glu Phe Lys Lys Glu Asn Gly Ile Asp Leu Ser Thr
 210 215 220
 Asp Lys Met Ala Met Gln Arg Leu Lys Asp Ala Ala Glu Lys Ala Lys
 225 230 235 240
 25 Lys Asp Leu Ser Gly Val Thr Ser Thr Gln Ile Ser Leu Pro Phe Ile
 245 250 255
 Thr Ala Gly Glu Ala Gly Pro Leu His Leu Glu Met Thr Leu Thr Arg
 260 265 270
 30 Ala Lys Phe Asp Asp Leu Thr Arg Asp Leu Val Glu Arg Thr Lys Val
 275 280 285
 35 Pro Val Arg Gln Ala Leu Ser Asp Ala Gly Leu Ser Leu Ser Glu Ile
 290 295 300
 Asp Glu Val Ile Leu Val Gly Gly Ser Thr Arg Ile Pro Ala Val Val
 305 310 315 320
 40 Glu Ala Val Lys Ala Glu Thr Gly Lys Glu Pro Asn Lys Ser Val Asn
 325 330 335
 Pro Asp Glu Val Val Ala Met Gly Ala Ala Ile Gln Gly Gly Val Ile
 340 345 350
 45 Thr Gly Asp Val Lys Asp Val Val Leu Leu Asp Val Thr Pro Leu Ser
 355 360 365
 50 Leu Gly Ile Glu Thr Met Gly Gly Val Phe Thr Lys Leu Ile Asp Arg
 370 375 380
 Asn Thr Thr Ile Pro Thr Ser Lys Ser Gln Val Phe Ser Thr Ala Ala
 385 390 395 400
 55 Asp Asn Gln Pro Ala Val Asp Ile His Val Leu Gln Gly Glu Arg Pro
 405 410 415
 Met Ala Ala Asp Asn Lys Thr Leu Gly Arg Phe Gln Leu Thr Asp Ile
 420 425 430
 60 Pro Ala Ala Pro Arg Gly Ile Pro Gln Ile Glu Val Thr Phe Asp Ile
 435 440 445
 65 Asp Lys Asn Gly Ile Val Ser Val Lys Ala Lys Asp Leu Gly Thr Gln
 450 455 460
 Lys Glu Gln Thr Ile Val Ile Gln Ser Asn Ser Gly Leu Thr Asp Glu
 465 470 475 480

Glu Ile Asp Arg Met Met Lys Asp Ala Glu Ala Asn Ala Glu Ser Asp
 485 490 495
 5 Lys Lys Arg Lys Glu Glu Val Asp Leu Arg Asn Glu Val Asp Gln Ala
 500 505 510
 Ile Phe Ala Thr Glu Lys Thr Ile Lys Glu Thr Glu Gly Lys Gly Phe
 515 520 525
 10 Asp Ala Glu Arg Asp Ala Ala Gln Ala Ala Leu Asp Asp Leu Lys Lys
 530 535 540
 Ala Gln Glu Asp Asn Asn Leu Asp Asp Met Lys Ala Lys Leu Glu Ala
 545 550 555 560
 Leu Asn Glu Lys Ala Gln Gly Leu Ala Val Lys Leu Tyr Glu Gln Ala
 565 570 575
 20 Ala Ala Ala Gln Gln Ala Gln Glu Gly Ala Glu Gly Ala Gln Ala Thr
 580 585 590
 Gly Asn Ala Gly Asp Asp Val Val Asp Gly Glu Phe Thr Glu Lys
 595 600 605
 25

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 352 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Asn Asn Thr Glu Phe Tyr Asp Arg Leu Gly Val Ser Lys Asn Ala
 1 5 10 15
 40 Ser Ala Asp Glu Ile Lys Lys Ala Tyr Arg Lys Leu Ser Lys Lys Tyr
 20 25 30
 45 His Pro Asp Ile Asn Lys Glu Pro Gly Ala Glu Asp Lys Tyr Lys Glu
 35 40 45
 Val Gln Glu Ala Tyr Glu Thr Leu Ser Asp Asp Gln Lys Arg Ala Ala
 50 55 60
 50 Tyr Asp Gln Tyr Gly Ala Ala Gly Ala Asn Gly Gly Phe Gly Gly Ala
 65 70 75 80
 Gly Gly Phe Gly Gly Phe Asn Gly Ala Gly Gly Phe Gly Gly Phe Glu
 85 90 95
 55 Asp Ile Phe Ser Ser Phe Phe Gly Gly Gly Gly Ser Ser Arg Asn Pro
 100 105 110
 60 Asn Ala Pro Arg Gln Gly Asp Asp Leu Gln Tyr Arg Val Asn Leu Thr
 115 120 125
 Phe Glu Glu Ala Ile Phe Gly Thr Glu Lys Glu Val Lys Tyr His Arg
 130 135 140
 65 Glu Ala Gly Cys Arg Thr Cys Asn Gly Ser Gly Ala Lys Pro Gly Thr
 145 150 155 160
 Ser Pro Val Thr Cys Gly Arg Cys His Gly Ala Gly Val Ile Asn Val
 165 170 175

Asp Thr Gln Thr Pro Leu Gly Met Met Arg Arg Gln Val Thr Cys Asp
 180 185 190

5 Val Cys His Gly Arg Gly Lys Glu Ile Lys Tyr Pro Cys Thr Thr Cys
 195 200 205

His Gly Thr Gly His Glu Lys Gln Ala His Ser Val His Val Lys Ile
 210 215 220

10 Pro Ala Gly Val Glu Thr Gly Gln Gln Ile Arg Leu Ala Gly Gln Gly
 225 230 235 240

Glu Ala Gly Phe Asn Gly Gly Pro Tyr Gly Asp Leu Tyr Val Val Val
 245 250 255

15 Ser Val Glu Ala Ser Asp Lys Phe Glu Arg Glu Gly Thr Thr Ile Phe
 260 265 270

20 Tyr Asn Leu Asn Leu Asn Phe Val Gln Ala Ala Leu Gly Asp Thr Val
 275 280 285

Asp Ile Pro Thr Val His Gly Asp Val Glu Leu Val Ile Pro Glu Gly
 290 295 300

25 Thr Gln Thr Gly Lys Lys Phe Arg Leu Arg Ser Lys Gly Ala Pro Ser
 305 310 315 320

Leu Arg Gly Gly Ala Val Gly Asp Gln Tyr Val Thr Val Asn Val Val
 325 330 335

30 Thr Pro Thr Gly Leu Asn Asp Arg Gln Lys Val Ala Leu Lys Glu Phe
 340 345 350

35

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: peptide

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Thr Ser Thr Gln Ile Ser Leu Pro Phe Ile Thr Ala Gly Glu Ala
 1 5 10 15

50

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: peptide

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Thr Ala Gly Glu Ala Gly Pro Leu His Leu Glu Met Thr Leu Thr
 1 5 10 15

65

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids

70

(B) TYPE: amino acid
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

10 Met Thr Leu Thr Arg Ala Lys Phe Asp Asp Leu Thr Arg Asp
1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

25 Asp Asp Leu Thr Arg Asp Leu Val Glu Arg Thr Lys Val Pro Val
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:11:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: peptide

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Thr Lys Val Pro Val Arg Gln Ala Leu Ser Asp Ala Gly Leu
1 5 10

45 (2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: peptide

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Lys Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile Val Ile
1 5 10 15

60 (2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

65 (ii) MOLECULE TYPE: peptide

70

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Leu Thr Asp Glu Ile Asp Arg Met Met Lys Asp Ala Glu Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Lys Asp Ala Glu Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu Glu
1 5 10 15

Val Asp Leu Arg Asn Glu Val Asp
20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr Ile Lys
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Glu Lys Thr Ile Lys Glu Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg
1 5 10 15

Asp Ala Ala Gln Ala Ala Leu Asp Asp Leu Lys Lys
20 25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Lys Ala Gln Glu Asp Asn Asn Leu Asp Asp Met Lys Ala Lys Leu Glu
1 5 10 15

5 Ala Leu Asn Glu Lys Ala Gln Gly Leu Ala Val Lys Leu Tyr
20 25 30

(2) INFORMATION FOR SEQ ID NO:18:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

20 Gln Glu Gly Ala Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp Asp
1 5 10 15
25 Val Val Asp Gly Glu Phe Thr Glu Lys
20 25

(2) INFORMATION FOR SEQ ID NO:19:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2183 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

40 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Streptococcus pyogenes

45 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 204..2030

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

50 CAGCGATGGT AGTTGTTTAT AACTAAGGTA AATGAGTTTT CGTTTTTGTC CGTAATGACA 60
GTAAACTAGA TAGCAAGTTA GAAGCTATTT CGCTTGCTGA TTAAACTATA GTGATTGCTT 120
55 AGAATTGGAA GTAAAATAAT TCGAGTGCTT ACTAAGATAA ATTGAAATAA AAAGTAATAA 180
AGTATAAAAT AAGAGGTATT AAC ATG TCT AAA ATT ATT GGT ATT GAC TTA 230
Met Ser Lys Ile Ile Gly Ile Asp Leu
1 5
60 GGT ACA ACA AAC TCA GCA GTA GCA GTT CTT GAA GGG ACT GAA TCA AAA 278
Gly Thr Thr Asn Ser Ala Val Ala Val Leu Glu Gly Thr Glu Ser Lys
10 15 20 25
65 ATC ATT GCT AAC CCA GAA GGC AAT CGT ACA ACT CCT TCA GTA GTA TCA 326
Ile Ile Ala Asn Pro Glu Gly Asn Arg Thr Thr Pro Ser Val Val Ser
30 35 40
70 TTC AAA AAT GGT GAA ATT ATC GTG GGT GAT GCT GCA AAA CGC CAA GCA 374
Phe Lys Asn Gly Glu Ile Ile Val Gly Asp Ala Ala Lys Arg Gln Ala
45 50 55

	GTG	ACA	AAC	CCA	GAA	ACA	GTA	ATC	TCT	ATT	AAA	TCT	AAA	ATG	GGA	ACT	422	-
	Val	Thr	Asn	Pro	Glu	Thr	Val	Ile	Ser	Ile	Lys	Ser	Lys	Met	Gly	Thr		
			60					65					70					
5	TCT	GAA	AAA	GTT	TCT	GCA	AAT	GGT	AAA	GAA	TAT	ACT	CCT	CAA	GAA	ATT	470	
	Ser	Glu	Lys	Val	Ser	Ala	Asn	Gly	Lys	Glu	Tyr	Thr	Pro	Gln	Glu	Ile		
		75					80					85						
10	TCA	GCA	ATG	ATT	CTT	CAA	TAC	CTT	AAA	GGT	TAT	GCT	GAA	GAC	TAT	CTT	518	
	Ser	Ala	Met	Ile	Leu	Gln	Tyr	Leu	Lys	Gly	Tyr	Ala	Glu	Asp	Tyr	Leu		
		90				95					100					105		
15	GGA	GAA	AAA	GTA	GAA	AAA	GCA	GTT	ATT	ACT	GTT	CCA	GCT	TAT	TTC	AAC	566	
	Gly	Glu	Lys	Val	Glu	Lys	Ala	Val	Ile	Thr	Val	Pro	Ala	Tyr	Phe	Asn		
				110						115					120			
20	GAT	GCA	CAA	CGT	CAA	GCA	ACT	AAA	GAC	GCT	GGT	AAA	ATT	GCA	GGT	CTT	614	
	Asp	Ala	Gln	Arg	Gln	Ala	Thr	Lys	Asp	Ala	Gly	Lys	Ile	Ala	Gly	Leu		
				125					130					135				
25	GAA	GTA	GAA	CGT	ATC	GTT	AAT	GAA	CCA	ACA	GCA	GCT	GCA	CTT	GCT	TAT	662	
	Glu	Val	Glu	Arg	Ile	Val	Asn	Glu	Pro	Thr	Ala	Ala	Ala	Leu	Ala	Tyr		
			140					145					150					
30	GGT	ATG	GAC	AAG	ACT	GAC	AAG	GAT	GAA	AAA	ATC	TTA	GTT	TTT	GAC	CTT	710	
	Gly	Met	Asp	Lys	Thr	Asp	Lys	Asp	Glu	Lys	Ile	Leu	Val	Phe	Asp	Leu		
		155					160					165						
35	GGT	GGT	GGT	ACA	TTT	GAC	GTA	TCA	ATC	CTT	GAA	TTA	GGT	GAT	GGT	GTC	758	
	Gly	Gly	Gly	Thr	Phe	Asp	Val	Ser	Ile	Leu	Glu	Leu	Gly	Asp	Gly	Val		
		170				175					180					185		
40	TTC	GAC	GTT	CTT	GCA	ACA	GCA	GGT	GAT	AAC	AAA	CTT	GGT	GGT	GAC	GAC	806	
	Phe	Asp	Val	Leu	Ala	Thr	Ala	Gly	Asp	Asn	Lys	Leu	Gly	Gly	Asp	Asp		
				190					195						200			
45	TTT	GAC	CAA	AAA	ATT	ATT	GAT	TTC	TTA	GTG	GCT	GAA	TTT	AAG	AAA	GAA	854	
	Phe	Asp	Gln	Lys	Ile	Ile	Asp	Phe	Leu	Val	Ala	Glu	Phe	Lys	Lys	Glu		
			205					210						215				
50	AAT	GGT	ATT	GAC	TTA	TCA	CAA	GAT	AAG	ATG	GCA	CTT	CAA	CGC	TTG	AAA	902	
	Asn	Gly	Ile	Asp	Leu	Ser	Gln	Asp	Lys	Met	Ala	Leu	Gln	Arg	Leu	Lys		
			220					225					230					
55	GAT	GCT	GCT	GAA	AAA	GCT	AAA	AAA	GAT	CTT	TCA	GGT	GTG	ACA	CAA	ACA	950	
	Asp	Ala	Ala	Glu	Lys	Ala	Lys	Lys	Asp	Leu	Ser	Gly	Val	Thr	Gln	Thr		
		235				240						245						
60	CAA	ATT	TCA	TTA	CCG	TTC	ATC	ACT	GCT	GGT	TCT	GCT	GGT	CCT	CTT	CAC	998	
	Gln	Ile	Ser	Leu	Pro	Phe	Ile	Thr	Ala	Gly	Ser	Ala	Gly	Pro	Leu	His		
		250				255					260					265		
65	TTA	GAG	ATG	AGC	TTA	TCT	CGT	GCT	AAA	TTT	GAC	GAT	CTC	ACT	CGT	GAC	1046	
	Leu	Glu	Met	Ser	Leu	Ser	Arg	Ala	Lys	Phe	Asp	Asp	Leu	Thr	Arg	Asp		
				270						275					280			
70	CTT	GTT	GAA	CGT	ACG	AAA	ACT	CCA	GTT	CGT	CAA	GCT	CTT	TCA	GAT	GCA	1094	
	Leu	Val	Glu	Arg	Thr	Lys	Thr	Pro	Val	Arg	Gln	Ala	Leu	Ser	Asp	Ala		
			285					290						295				
75	GGA	TTG	TCA	TTG	TCA	GAA	ATT	GAT	GAA	GTT	ATC	CTT	GTT	GGT	GGA	TCA	1142	
	Gly	Leu	Ser	Leu	Ser	Glu	Ile	Asp	Glu	Val	Ile	Leu	Val	Gly	Gly	Ser		
			300					305					310					
80	ACT	CGT	ATC	CCA	GCA	GTT	GTC	GAA	GCT	GTA	AAA	GCT	GAA	ACT	GGT	AAA	1190	
	Thr	Arg	Ile	Pro	Ala	Val	Val	Glu	Ala	Val	Lys	Ala	Glu	Thr	Gly	Lys		
		315					320					325						

5	GAA CCA AAT AAA TCT GTA AAC CCT GAT GAA GTG GTT GCT ATG GGT GCT Glu Pro Asn Lys Ser Val Asn Pro Asp Glu Val Val Ala Met Gly Ala 330 335 340 345	1238	—
10	GCT ATC CAA GGT GGG GTT ATC ACT GGG GAT GTG AAA GAC GTT GTC CTT Ala Ile Gln Gly Gly Val Ile Thr Gly Asp Val Lys Asp Val Val Leu 350 355 360	1286	
15	CTT GAC GTA ACA CCA TTG TCA CTT GGT ATT GAA ACA ATG GGT GGT GTC Leu Asp Val Thr Pro Leu Ser Leu Gly Ile Glu Thr Met Gly Gly Val 365 370 375	1334	
20	TTC ACT AAA TTG ATC GAC CGC AAT ACA ACT ATC CCA ACA TCT AAA TCA Phe Thr Lys Leu Ile Asp Arg Asn Thr Thr Ile Pro Thr Ser Lys Ser 380 385 390	1382	
25	CAA GTC TTC TCA ACA GCA GCA GAC AAC CAA CCA GCC GTT GAT ATC CAT Gln Val Phe Ser Thr Ala Ala Asp Asn Gln Pro Ala Val Asp Ile His 395 400 405	1430	
30	GTT CTT CAA GGT GAA CGC CCA ATG GCA GCA GAT AAC AAG ACT CTT GGT Val Leu Gln Gly Glu Arg Pro Met Ala Ala Asp Asn Lys Thr Leu Gly 410 415 420 425	1478	
35	CGC TTC CAA TTG ACT GAT ATC CCA GCT GCA CCT CGT GGA ATC CCA CAA Arg Phe Gln Leu Thr Asp Ile Pro Ala Ala Pro Arg Gly Ile Pro Gln 430 435 440	1526	
40	ATT GAA GTA ACA TTT GAT ATC GAT AAA AAC GGT ATT GTT TCT GTA AAA Ile Glu Val Thr Phe Asp Ile Asp Lys Asn Gly Ile Val Ser Val Lys 445 450 455	1574	
45	GCT AAA GAC CTT GGT ACG CAA AAG GAA CAA CAC ATC GTT ATC AAA TCA Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln His Ile Val Ile Lys Ser 460 465 470	1622	
50	AAC GAC GGA CTT TCT GAA GAA GAA ATT GAT CGC ATG ATG AAA GAC GCT Asn Asp Gly Leu Ser Glu Glu Glu Ile Asp Arg Met Met Lys Asp Ala 475 480 485	1670	
55	GAA GCT AAT GCC GAA GCC GAT GCG AAA CGT AAA GAA GAA GTT GAC CTT Glu Ala Asn Ala Glu Ala Asp Ala Lys Arg Lys Glu Glu Val Asp Leu 490 495 500 505	1718	
60	AAA AAC GAA GTT GAC CAA GCT ATC TTT GCT ACT GAA AAA ACA ATC AAA Lys Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr Ile Lys 510 515 520	1766	
65	GAA ACT GAA GGT AAA GGC TTT GAC ACA GAA CGC GAT GCA GCG CAA TCA Glu Thr Glu Gly Lys Gly Phe Asp Thr Glu Arg Asp Ala Ala Gln Ser 525 530 535	1814	
70	GCT CTT GAC GAG TTA AAA GCT GCG CAA GAA TCT GGC AAC CTT GAC GAC Ala Leu Asp Glu Leu Lys Ala Ala Gln Glu Ser Gly Asn Leu Asp Asp 540 545 550	1862	
75	ATG AAA GCT AAA CTT GAA GCA TTA AAT GAA AAA GCG CAA GCT TTG GCT Met Lys Ala Lys Leu Glu Ala Leu Asn Glu Lys Ala Gln Ala Leu Ala 555 560 565	1910	
80	GTT AAA ATG TAC GAG CAA GCT GCA GCA GCT CAA CAA GCA GCA CAA GGT Val Lys Met Tyr Glu Gln Ala Ala Ala Gln Ala Ala Gln Gly 570 575 580 585	1958	
85	GCA GAA GGT GCA CAA GCT AAT GAT TCA GCA AAT AAT GAT GAT GTT GTA Ala Glu Gly Ala Gln Ala Asn Asp Ser Ala Asn Asn Asp Asp Val Val 590 595 600	2006	

GAT GGC GAA TTT ACA GAA AAG TAATGATTTA GTTATCTAGT AACATTAATA
 Asp Gly Glu Phe Thr Glu Lys
 605

2057 -

5

TCCGAATTCA GAGGTTGTAC CAAACCTCTG TTTTGGCTA AATAAAATGT AAAAATGCTG

2117

ACGTCAAAAT ATTTTAAGAA AGGAATACAA GTTCGATTAT TCGAACACAG GCTAAAGCGT

2177

10 GTAAAG

2183

(2) INFORMATION FOR SEQ ID NO:20:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 608 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

25	Met	Ser	Lys	Ile	Ile	Gly	Ile	Asp	Leu	Gly	Thr	Thr	Asn	Ser	Ala	Val	1	5	10	15
	Ala	Val	Leu	Glu	Gly	Thr	Glu	Ser	Lys	Ile	Ile	Ala	Asn	Pro	Glu	Gly	20	25	30	
30	Asn	Arg	Thr	Thr	Pro	Ser	Val	Val	Ser	Phe	Lys	Asn	Gly	Glu	Ile	Ile	35	40	45	
	Val	Gly	Asp	Ala	Ala	Lys	Arg	Gln	Ala	Val	Thr	Asn	Pro	Glu	Thr	Val	50	55	60	
35	Ile	Ser	Ile	Lys	Ser	Lys	Met	Gly	Thr	Ser	Glu	Lys	Val	Ser	Ala	Asn	65	70	75	80
40	Gly	Lys	Glu	Tyr	Thr	Pro	Gln	Glu	Ile	Ser	Ala	Met	Ile	Leu	Gln	Tyr	85	90	95	
	Leu	Lys	Gly	Tyr	Ala	Glu	Asp	Tyr	Leu	Gly	Glu	Lys	Val	Glu	Lys	Ala	100	105	110	
45	Val	Ile	Thr	Val	Pro	Ala	Tyr	Phe	Asn	Asp	Ala	Gln	Arg	Gln	Ala	Thr	115	120	125	
	Lys	Asp	Ala	Gly	Lys	Ile	Ala	Gly	Leu	Glu	Val	Glu	Arg	Ile	Val	Asn	130	135	140	
50	Glu	Pro	Thr	Ala	Ala	Ala	Leu	Ala	Tyr	Gly	Met	Asp	Lys	Thr	Asp	Lys	145	150	155	160
55	Asp	Glu	Lys	Ile	Leu	Val	Phe	Asp	Leu	Gly	Gly	Gly	Thr	Phe	Asp	Val	165	170	175	
	Ser	Ile	Leu	Glu	Leu	Gly	Asp	Gly	Val	Phe	Asp	Val	Leu	Ala	Thr	Ala	180	185	190	
60	Gly	Asp	Asn	Lys	Leu	Gly	Gly	Asp	Asp	Phe	Asp	Gln	Lys	Ile	Ile	Asp	195	200	205	
	Phe	Leu	Val	Ala	Glu	Phe	Lys	Lys	Glu	Asn	Gly	Ile	Asp	Leu	Ser	Gln	210	215	220	
65	Asp	Lys	Met	Ala	Leu	Gln	Arg	Leu	Lys	Asp	Ala	Ala	Glu	Lys	Ala	Lys	225	230	235	240
70	Lys	Asp	Leu	Ser	Gly	Val	Thr	Gln	Thr	Gln	Ile	Ser	Leu	Pro	Phe	Ile	245	250	255	

Thr Ala Gly Ser Ala Gly Pro Leu His Leu Glu Met Ser Leu Ser Arg
 260 265 270
 5 Ala Lys Phe Asp Asp Leu Thr Arg Asp Leu Val Glu Arg Thr Lys Thr
 275 280 285
 Pro Val Arg Gln Ala Leu Ser Asp Ala Gly Leu Ser Leu Ser Glu Ile
 290 295 300
 10 Asp Glu Val Ile Leu Val Gly Gly Ser Thr Arg Ile Pro Ala Val Val
 305 310 315 320
 Glu Ala Val Lys Ala Glu Thr Gly Lys Glu Pro Asn Lys Ser Val Asn
 325 330 335
 15 Pro Asp Glu Val Val Ala Met Gly Ala Ala Ile Gln Gly Gly Val Ile
 340 345 350
 20 Thr Gly Asp Val Lys Asp Val Val Leu Leu Asp Val Thr Pro Leu Ser
 355 360 365
 Leu Gly Ile Glu Thr Met Gly Gly Val Phe Thr Lys Leu Ile Asp Arg
 370 375 380
 25 Asn Thr Thr Ile Pro Thr Ser Lys Ser Gln Val Phe Ser Thr Ala Ala
 385 390 395 400
 Asp Asn Gln Pro Ala Val Asp Ile His Val Leu Gln Gly Glu Arg Pro
 405 410 415
 30 Met Ala Ala Asp Asn Lys Thr Leu Gly Arg Phe Gln Leu Thr Asp Ile
 420 425 430
 35 Pro Ala Ala Pro Arg Gly Ile Pro Gln Ile Glu Val Thr Phe Asp Ile
 435 440 445
 Asp Lys Asn Gly Ile Val Ser Val Lys Ala Lys Asp Leu Gly Thr Gln
 450 455 460
 40 Lys Glu Gln His Ile Val Ile Lys Ser Asn Asp Gly Leu Ser Glu Glu
 465 470 475 480
 Glu Ile Asp Arg Met Met Lys Asp Ala Glu Ala Asn Ala Glu Ala Asp
 485 490 495
 45 Ala Lys Arg Lys Glu Glu Val Asp Leu Lys Asn Glu Val Asp Gln Ala
 500 505 510
 50 Ile Phe Ala Thr Glu Lys Thr Ile Lys Glu Thr Glu Gly Lys Gly Phe
 515 520 525
 Asp Thr Glu Arg Asp Ala Ala Gln Ser Ala Leu Asp Glu Leu Lys Ala
 530 535 540
 55 Ala Gln Glu Ser Gly Asn Leu Asp Asp Met Lys Ala Lys Leu Glu Ala
 545 550 555 560
 Leu Asn Glu Lys Ala Gln Ala Leu Ala Val Lys Met Tyr Glu Gln Ala
 565 570 575
 60 Ala Ala Ala Gln Gln Ala Ala Gln Gly Ala Glu Gly Ala Gln Ala Asn
 580 585 590
 65 Asp Ser Ala Asn Asn Asp Asp Val Val Asp Gly Glu Phe Thr Glu Lys
 595 600 605

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2438 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: *Streptococcus agalactiae*

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 248..2077

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

25	CTTTCAAAAG GGATATAAAT TGCACGAGCG TCTGCTAAGA CCAGCGATGG TAGTTGTCTA	60
	TAACTAAGGT AAATGAGTTT TCGTTTTTGT CCGTAATGAC AGTAACTAG ATAGCAAGTT	120
30	AGAAGCTATT CAGCTTGCTG ATTAACTAT AGTGATTGCT TAGAATTGGA AGTAAAATAA	180
	TTCGAGTGCT TACTAAGATA AATTGAAATA AAAAGTAATA AAGTATTATA AAATAAGAGG	240
35	TATTAAC ATG TCT AAA ATT ATT GGT ATT GAC TTA GGT ACA ACA AAC TCA	289
	Met Ser Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser	
	1 5 10	
	GCA GTA GCA GTT CTT GAA GGG ACT GAA TCA AAA ATC ATT GCT AAC CCA	337
40	Ala Val Ala Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro	
	15 20 25 30	
	GAA GGC AAT CGT ACA ACT CCT TCA GTA GTA TCA TTC AAA AAT GGT GAA	385
	Glu Gly Asn Arg Thr Thr Pro Ser Val Val Ser Phe Lys Asn Gly Glu	
	35 40 45	
45	ATT ATC GTG GGT GAT GCT GCA AAA CGT CAA GCG GTA ACA AAT CCA GAT	433
	Ile Ile Val Gly Asp Ala Ala Lys Arg Gln Ala Val Thr Asn Pro Asp	
	50 55 60	
50	ACT GTT ATC TCT ATC AAA TCA AAG ATG GGA ACT TCT GAA AAA GTT TCT	481
	Thr Val Ile Ser Ile Lys Ser Lys Met Gly Thr Ser Glu Lys Val Ser	
	65 70 75	
55	GCA AAT GGT AAA GAA TAT ACT CCT CAA GAA ATT TCA GCA ATG ATT CTT	529
	Ala Asn Gly Lys Glu Tyr Thr Pro Gln Glu Ile Ser Ala Met Ile Leu	
	80 85 90	
	CAA TAC CTT AAA GGT TAT GCT GAA GAC TAT CTT GGA GAA AAA GTA GAA	577
60	Gln Tyr Leu Lys Gly Tyr Ala Glu Asp Tyr Leu Gly Glu Lys Val Glu	
	95 100 105 110	
	AAA GCA GTT ATT ACT GTT CCA GCT TAC TTC AAC GAT GCA CAA CGT CAG	625
	Lys Ala Val Ile Thr Val Pro Ala Tyr Phe Asn Asp Ala Gln Arg Gln	
	115 120 125	
65	GCA ACT AAA GAC GCT GGT AAA ATT GCA GGT CTT GAA GTA GAA CGT ATC	673
	Ala Thr Lys Asp Ala Gly Lys Ile Ala Gly Leu Glu Val Glu Arg Ile	
	130 135 140	

	GTT AAC GAA CCA ACA GCA GCC GCA CTT GCT TAT GGT ATG GAC AAG ACT	721	-
	Val Asn Glu Pro Thr Ala Ala Ala Leu Ala Tyr Gly Met Asp Lys Thr		
	145 150 155		
5	GAC AAG GAT GAA AAA ATC TTA GTT TTT GAC CTT GGT GGT GGT ACA TTT	769	
	Asp Lys Asp Glu Lys Ile Leu Val Phe Asp Leu Gly Gly Gly Thr Phe		
	160 165 170		
10	GAC GTA TCA ATC CTT GAA TTA GGT GAT GGT GTC TTC GAC GTT CTT GCA	817	
	Asp Val Ser Ile Leu Glu Leu Gly Asp Gly Val Phe Asp Val Leu Ala		
	175 180 185 190		
15	ACA GCA GGT GAT AAC AAA CTT GGT GGT GAC GAC TTT GAC CAG AAA ATT	865	
	Thr Ala Gly Asp Asn Lys Leu Gly Gly Asp Asp Phe Asp Gln Lys Ile		
	195 200 205		
20	ATT GAT TTC TTG GTA GAA GAA TTC AAG AAA GAA AAT GGT ATT GAT CTT	913	
	Ile Asp Phe Leu Val Glu Glu Phe Lys Lys Glu Asn Gly Ile Asp Leu		
	210 215 220		
25	TCT CAA GAC AAA ATG GCT CTT CAA CGC TTG AAA GAT GCT GCT GAA AAA	961	
	Ser Gln Asp Lys Met Ala Leu Gln Arg Leu Lys Asp Ala Ala Glu Lys		
	225 230 235		
30	GCT AAA AAA GAC CTT TCA GGT GTA ACT CAA ACT CAA ATT TCA TTA CCG	1009	
	Ala Lys Lys Asp Leu Ser Gly Val Thr Gln Thr Gln Ile Ser Leu Pro		
	240 245 250		
35	TTC ATC ACT GCT GGT TCT GCT GGT CCT CTT CAC TTG GAG ATG AGC TTA	1057	
	Phe Ile Thr Ala Gly Ser Ala Gly Pro Leu His Leu Glu Met Ser Leu		
	255 260 265 270		
40	TCA CGT GCT AAA TTT GAC GAT CTC ACT CGT GAC CTT GTT GAA CGT ACG	1105	
	Ser Arg Ala Lys Phe Asp Asp Leu Thr Arg Asp Leu Val Glu Arg Thr		
	275 280 285		
45	AAA ACT CCA GTT CGT CAA GCT CTT TCA GAT GCA GGC TTG TCA TTG TCA	1153	
	Lys Thr Pro Val Arg Gln Ala Leu Ser Asp Ala Gly Leu Ser Leu Ser		
	290 295 300		
50	GAA ATT GAT GAA GTT ATC CTC GTT GGT GGA TCA ACA CGT ATC CCA GCA	1201	
	Glu Ile Asp Glu Val Ile Leu Val Gly Gly Ser Thr Arg Ile Pro Ala		
	305 310 315		
55	GTT GTT GAA GCT GTA AAA GCT GAA ACT GGT AAA GAA CCA AAT AAA TCT	1249	
	Val Val Glu Ala Val Lys Ala Glu Thr Gly Lys Glu Pro Asn Lys Ser		
	320 325 330		
60	GTT AAC CCT GAT GAA GTG GTT GCC ATG GGT GCT GCT ATC CAA GGT GGT	1297	
	Val Asn Pro Asp Glu Val Val Ala Met Gly Ala Ala Ile Gln Gly Gly		
	335 340 345 350		
65	GTT ATC ACT GGG GAT GTG AAA GAC GTT GTA CTT CTT GAC GTA ACA CCA	1345	
	Val Ile Thr Gly Asp Val Lys Asp Val Val Leu Leu Asp Val Thr Pro		
	355 360 365		
70	TTG TCA CTT GGT ATT GAA ACA ATG GGT GGT GTC TTC ACT AAA TTG ATC	1393	
	Leu Ser Leu Gly Ile Glu Thr Met Gly Gly Val Phe Thr Lys Leu Ile		
	370 375 380		
75	GAC CGC AAC ACA ACT ATC CCA ACA TCT AAA TCA CAA GTC TTC TCA ACA	1441	
	Asp Arg Asn Thr Thr Ile Pro Thr Ser Lys Ser Gln Val Phe Ser Thr		
	385 390 395		
80	GCA GCA GAC AAC CAA CCA GCC GTT GAT ATC CAT GTT CTT CAA GGT GAA	1489	
	Ala Ala Asp Asn Gln Pro Ala Val Asp Ile His Val Leu Gln Gly Glu		
	400 405 410		

	CGC CCA ATG GCA GCA GAT AAC AAA ACA CTC GGT CGC TTC CAA TTG ACT	1537	-
	Arg Pro Met Ala Ala Asp Asn Lys Thr Leu Gly Arg Phe Gln Leu Thr		
	415 420 425 430		
5	GAT ATC CCA GCT GCA CCT CGT GGA ATC CCA CAA ATT GAA GTA ACA TTT	1585	
	Asp Ile Pro Ala Ala Pro Arg Gly Ile Pro Gln Ile Glu Val Thr Phe		
	435 440 445		
10	GAT ATC GAT AAA AAT GGT ATT GTA TCT GTT AAA GCT AAA GAT CTC GGT	1633	
	Asp Ile Asp Lys Asn Gly Ile Val Ser Val Lys Ala Lys Asp Leu Gly		
	450 455 460		
15	ACT CAA AAA GAA CAA CAC ATT GTT ATC CAA TCT AAT TCA GGA TTA ACT	1681	
	Thr Gln Lys Glu Gln His Ile Val Ile Gln Ser Asn Ser Gly Leu Thr		
	465 470 475		
20	GAT GAA GAA ATT GAT AAA ATG ATG AAA GAT GCT GAA GCA AAT GCT GAA	1729	
	Asp Glu Glu Ile Asp Lys Met Met Lys Asp Ala Glu Ala Asn Ala Glu		
	480 485 490		
25	GCA GAT GCA AAA CGT AAA GAA GAA GTT GAT CTT AAA AAT GAA GTT GAC	1777	
	Ala Asp Ala Lys Arg Lys Glu Glu Val Asp Leu Lys Asn Glu Val Asp		
	495 500 505 510		
	CAA GCC ATC TTT GCA ACA GAA AAA ACT ATT AAA GAA ACT GAA GGC AAA	1825	
	Gln Ala Ile Phe Ala Thr Glu Lys Thr Ile Lys Glu Thr Glu Gly Lys		
	515 520 525		
30	GGT TTT GAT ACA GAA CGC GAT GCA GCG CAA TCA GCA CTT GAT GAG TTG	1873	
	Gly Phe Asp Thr Glu Arg Asp Ala Ala Gln Ser Ala Leu Asp Glu Leu		
	530 535 540		
35	AAA AAA GCT CAA GAA TCA GGT AAC CTT GAC GAC ATG AAA GCT AAA CTT	1921	
	Lys Lys Ala Gln Glu Ser Gly Asn Leu Asp Asp Met Lys Ala Lys Leu		
	545 550 555		
40	GAA GCT CTT AAC GAA AAA GCA CAA GCT CTT GCA GTT AAA CTT TAC GAA	1969	
	Glu Ala Leu Asn Glu Lys Ala Gln Ala Leu Ala Val Lys Leu Tyr Glu		
	560 565 570		
	CAA GCG GCT GCA GCA CAA CAA GCA GCT CAA GGG GCT GAA GGT GCA CAA	2017	
	Gln Ala Ala Ala Ala Gln Gln Ala Ala Gln Gly Ala Glu Gly Ala Gln		
	575 580 585 590		
45	TCA GCT GAT TCA TCA AGC AAG GGT GAT GAT GTT GTA GAT GGC GAA TTC	2065	
	Ser Ala Asp Ser Ser Ser Lys Gly Asp Val Val Asp Gly Glu Phe		
	595 600 605		
50	ACT GAG AAA TAATTATTAA TATTGTTTCAG ATTCAATTGA ATATAAGCAT	2114	
	Thr Glu Lys		
	610		
55	GAAACTATA CTAGCATAGT AAAGTTCTTC GTGATAGGGA TTGCTCAATA ATCTAGATAA	2174	
	GTTCAGATT ACATAAGCTA ATTCGCTAT CACTAAATAA AAACATATTA ATAATAAATA	2234	
	GGCGGGGCGC CTCGCTCCGT CTGTTTTATT AAGTGTCATA TATATGTTAA CTATTTAGAG	2294	
60	CTGTAAGTGG GCAAGAATAA TTGTTAATCT CTTCAAGTGT AGTATATGAA CAAAATATAA	2354	
	AGGATTAGAT AATGAACAAT ACAGAAATTTT ATGATCGTCT TGGCGTTTCA AAAGATGCTT	2414	
	CTCAGGACGA AATAAAAAAA GCTT	2438	
65			

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- 70 (A) LENGTH: 609 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```

Met Ser Lys Ile Ile Gly Ile Asp Leu Gly Thr Thr Asn Ser Ala Val
 1           5           10           15
10 Ala Val Leu Glu Gly Thr Glu Ser Lys Ile Ile Ala Asn Pro Glu Gly
           20           25           30
    Asn Arg Thr Thr Pro Ser Val Val Ser Phe Lys Asn Gly Glu Ile Ile
           35           40           45
15 Val Gly Asp Ala Ala Lys Arg Gln Ala Val Thr Asn Pro Asp Thr Val
           50           55           60
    Ile Ser Ile Lys Ser Lys Met Gly Thr Ser Glu Lys Val Ser Ala Asn
20           65           70           75           80
    Gly Lys Glu Tyr Thr Pro Gln Glu Ile Ser Ala Met Ile Leu Gln Tyr
           85           90           95
25 Leu Lys Gly Tyr Ala Glu Asp Tyr Leu Gly Glu Lys Val Glu Lys Ala
           100          105          110
    Val Ile Thr Val Pro Ala Tyr Phe Asn Asp Ala Gln Arg Gln Ala Thr
           115          120          125
30 Lys Asp Ala Gly Lys Ile Ala Gly Leu Glu Val Glu Arg Ile Val Asn
           130          135          140
    Glu Pro Thr Ala Ala Ala Leu Ala Tyr Gly Met Asp Lys Thr Asp Lys
35           145          150          155          160
    Asp Glu Lys Ile Leu Val Phe Asp Leu Gly Gly Gly Thr Phe Asp Val
           165          170          175
40 Ser Ile Leu Glu Leu Gly Asp Gly Val Phe Asp Val Leu Ala Thr Ala
           180          185          190
    Gly Asp Asn Lys Leu Gly Gly Asp Asp Phe Asp Gln Lys Ile Ile Asp
           195          200          205
45 Phe Leu Val Glu Glu Phe Lys Lys Glu Asn Gly Ile Asp Leu Ser Gln
           210          215          220
    Asp Lys Met Ala Leu Gln Arg Leu Lys Asp Ala Ala Glu Lys Ala Lys
50           225          230          235          240
    Lys Asp Leu Ser Gly Val Thr Gln Thr Gln Ile Ser Leu Pro Phe Ile
           245          250          255
55 Thr Ala Gly Ser Ala Gly Pro Leu His Leu Glu Met Ser Leu Ser Arg
           260          265          270
    Ala Lys Phe Asp Asp Leu Thr Arg Asp Leu Val Glu Arg Thr Lys Thr
           275          280          285
60 Pro Val Arg Gln Ala Leu Ser Asp Ala Gly Leu Ser Leu Ser Glu Ile
           290          295          300
    Asp Glu Val Ile Leu Val Gly Gly Ser Thr Arg Ile Pro Ala Val Val
65           305          310          315          320
    Glu Ala Val Lys Ala Glu Thr Gly Lys Glu Pro Asn Lys Ser Val Asn
           325          330          335
70 Pro Asp Glu Val Val Ala Met Gly Ala Ala Ile Gln Gly Gly Val Ile
           340          345          350

```

Thr Gly Asp Val Lys Asp Val Val Leu Leu Asp Val Thr Pro Leu Ser
 355 360 365
 5 Leu Gly Ile Glu Thr Met Gly Gly Val Phe Thr Lys Leu Ile Asp Arg
 370 375 380
 Asn Thr Thr Ile Pro Thr Ser Lys Ser Gln Val Phe Ser Thr Ala Ala
 385 390 395 400
 10 Asp Asn Gln Pro Ala Val Asp Ile His Val Leu Gln Gly Glu Arg Pro
 405 410 415
 Met Ala Ala Asp Asn Lys Thr Leu Gly Arg Phe Gln Leu Thr Asp Ile
 420 425 430
 Pro Ala Ala Pro Arg Gly Ile Pro Gln Ile Glu Val Thr Phe Asp Ile
 435 440 445
 20 Asp Lys Asn Gly Ile Val Ser Val Lys Ala Lys Asp Leu Gly Thr Gln
 450 455 460
 Lys Glu Gln His Ile Val Ile Gln Ser Asn Ser Gly Leu Thr Asp Glu
 465 470 475 480
 25 Glu Ile Asp Lys Met Met Lys Asp Ala Glu Ala Asn Ala Glu Ala Asp
 485 490 495
 Ala Lys Arg Lys Glu Glu Val Asp Leu Lys Asn Glu Val Asp Gln Ala
 500 505 510
 30 Ile Phe Ala Thr Glu Lys Thr Ile Lys Glu Thr Glu Gly Lys Gly Phe
 515 520 525
 35 Asp Thr Glu Arg Asp Ala Ala Gln Ser Ala Leu Asp Glu Leu Lys Lys
 530 535 540
 Ala Gln Glu Ser Gly Asn Leu Asp Asp Met Lys Ala Lys Leu Glu Ala
 545 550 555 560
 40 Leu Asn Glu Lys Ala Gln Ala Leu Ala Val Lys Leu Tyr Glu Gln Ala
 565 570 575
 Ala Ala Ala Gln Gln Ala Ala Gln Gly Ala Glu Gly Ala Gln Ser Ala
 580 585 590
 45 Asp Ser Ser Ser Lys Gly Asp Asp Val Val Asp Gly Glu Phe Thr Glu
 595 600 605
 50 Lys

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

Arg Ile Pro Ala Val Val Glu Ala Val Lys Ala Glu Thr Gly Lys Glu
 1 5 10 15

Pro Asn Lys

(2) INFORMATION FOR SEQ ID NO:24:

- 5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

15 Gln Thr Ile Val Ile Gln Ser Asn Ser Gly Leu Thr Asp Glu Glu
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:25:

- 20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 460 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

30 (iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: *Streptococcus pneumoniae*

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..456

(D) OTHER INFORMATION: /product= "C-terminal 151-residue
 fragment (C-151) of HSP72"

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

45	ATG AAG GCC AAA GAC CTT GGA ACT CAA AAA GAA CAA ACT ATT GTC ATC Met Lys Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile Val Ile 1 5 10 15	48
50	CAA TCG AAC TCA GGT TTG ACT GAC GAA GAA ATC GAC CGC ATG ATG AAA Gln Ser Asn Ser Gly Leu Thr Asp Glu Glu Ile Asp Arg Met Met Lys 20 25 30	96
55	GAT GCA GAA GCA AAC GCT GAA TCC GAT AAG AAA CGT AAA GAA GAA GTA Asp Ala Glu Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu Glu Val 35 40 45	144
60	GAC CTT CGT AAT GAA GTG GAC CAA GCA ATC TTT GCG ACT GAA AAG ACA Asp Leu Arg Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr 50 55 60	192
65	ATC AAG GAA ACT GAA GGT AAA GGC TTC GAC GCA GAA CGT GAC GCT GCC Ile Lys Glu Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg Asp Ala Ala 65 70 75 80	240
70	CAA GCT GCC CTT GAT GAC CTT AAG AAA GCT CAA GAA GAC AAC AAC TTG Gln Ala Ala Leu Asp Asp Leu Lys Lys Ala Gln Glu Asp Asn Asn Leu 85 90 95	288
75	GAC GAC ATG AAA GCA AAA CTT GAA GCA TTG AAC GAA AAA GCT CAA GGA Asp Asp Met Lys Ala Lys Leu Glu Ala Leu Asn Glu Lys Ala Gln Gly 100 105 110	336

CTT GCT GTT AAA CTC TAC GAA CAA GCC GCA GCA GCG CAA CAA GCT CAA 384
 Leu Ala Val Lys Leu Tyr Glu Gln Ala Ala Ala Gln Gln Ala Gln
 115 120 125

5 GAA GGA GCA GAA GGC GCA CAA GCA ACA GGA AAC GCA GGC GAT GAC GTC 432
 Glu Gly Ala Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp Asp Val
 130 135 140

10 GTA GAC GGA GAG TTT ACG GAA AAG TAAG 460
 Val Asp Gly Glu Phe Thr Glu Lys
 145 150

(2) INFORMATION FOR SEQ ID NO:26:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

25 Met Lys Ala Lys Asp Leu Gly Thr Gln Lys Glu Gln Thr Ile Val Ile
 1 5 10 15

Gln Ser Asn Ser Gly Leu Thr Asp Glu Ile Asp Arg Met Met Lys
 20 25 30

30 Asp Ala Glu Ala Asn Ala Glu Ser Asp Lys Lys Arg Lys Glu Glu Val
 35 40 45

35 Asp Leu Arg Asn Glu Val Asp Gln Ala Ile Phe Ala Thr Glu Lys Thr
 50 55 60

Ile Lys Glu Thr Glu Gly Lys Gly Phe Asp Ala Glu Arg Asp Ala Ala
 65 70 75 80

40 Gln Ala Ala Leu Asp Asp Leu Lys Lys Ala Gln Glu Asp Asn Asn Leu
 85 90 95

Asp Asp Met Lys Ala Lys Leu Glu Ala Leu Asn Glu Lys Ala Gln Gly
 100 105 110

45 Leu Ala Val Lys Leu Tyr Glu Gln Ala Ala Ala Ala Gln Gln Ala Gln
 115 120 125

50 Glu Gly Ala Glu Gly Ala Gln Ala Thr Gly Asn Ala Gly Asp Asp Val
 130 135 140

Val Asp Gly Glu Phe Thr Glu Lys
 145 150

We claim:

1. A polypeptide selected from the group consisting of:

- 5 (a) the HSP72 polypeptide having the amino acid sequence of SEQ ID NO:5;
- (b) the HSP70 (DnaK) polypeptide having the amino acid sequence of SEQ ID NO:20;
- (c) the HSP70 (DnaK) polypeptide having
10 the amino acid sequence of SEQ ID NO:22;
- (d) polypeptides that are immunologically reactive with antibodies generated by infection of a mammalian host with *Streptococcus pneumoniae* cells, which antibodies are immunologically reactive with the
15 polypeptide of paragraph (a), (b), or (c);
- (e) polypeptides that are capable of eliciting antibodies that are immunologically reactive with the polypeptide of paragraph (a), (b), or (c);
- (f) polypeptides that are immunologically
20 reactive with antibodies elicited by immunization with the polypeptide of paragraph (a), (b), or (c); and
- (g) fragments of any of the foregoing polypeptides, either alone or in combination with other polypeptides to form a fusion protein.

25

2. The polypeptide of claim 1, wherein the polypeptides of paragraph (d) are selected from the group consisting of polypeptides of the genera *Streptococcus* and *Enterococcus*.

30

3. The polypeptide of claim 1, wherein the polypeptides of paragraph (d) are selected from the group consisting of polypeptides of the species *Streptococcus pneumoniae*, *Streptococcus agalactiae*, *Streptococcus*

pyogenes, *Streptococcus mutans*, *Streptococcus sanguis*, and *Enterococcus faecalis*.

4. The polypeptide of claim 1, wherein the
5 polypeptides of paragraph (d) are selected from the group
consisting of polypeptides of the species *Streptococcus*
pneumoniae, *Streptococcus agalactiae*, and *Streptococcus*
pyogenes.

10 5. The polypeptide of claim 1, wherein the
fragments of paragraph (g) are selected from the group
consisting of amino acids 439-607 of SEQ ID NO:5 (C-169),
amino acids 457-607 of SEQ ID NO:5 (C-151), amino acids
527-541 of SEQ ID NO:5, and amino acids 586-600 of SEQ ID
15 NO:5.

6. A polypeptide having the amino acid
sequence of SEQ ID NO:5 analogues, homologues and
derivatives thereof.

20

7. A polypeptide having the amino acid
sequence of SEQ ID NO:20, analogues, homologues and
derivatives thereof.

25 8. A polypeptide having the amino acid
sequence of SEQ ID NO:22, analogues, homologues and
derivatives thereof.

9. A polypeptide having the amino acid
30 sequence of SEQ ID NO:26, analogues, homologues and
derivatives thereof.

10. A polypeptide having the amino acid sequence of SEQ ID NO:7, analogues, homologues and derivatives thereof.
- 5 11. A polypeptide having the amino acid sequence of SEQ ID NO:8, analogues, homologues and derivatives thereof.
- 10 12. A polypeptide having the amino acid sequence of SEQ ID NO:9, analogues, homologues and derivatives thereof.
- 15 13. A polypeptide having the amino acid sequence of SEQ ID NO:10, analogues, homologues and derivatives thereof.
- 20 14. A polypeptide having the amino acid sequence of SEQ ID NO:11, analogues, homologues and derivatives thereof.
- 25 15. A polypeptide having the amino acid sequence of SEQ ID NO:12, analogues, homologues and derivatives thereof.
- 30 16. A polypeptide having the amino acid sequence of SEQ ID NO:13, analogues, homologues and derivatives thereof.
17. A polypeptide having the amino acid sequence of SEQ ID NO:14, analogues, homologues and derivatives thereof.

18. A polypeptide having the amino acid sequence of SEQ ID NO:15, analogues, homologues and derivatives thereof.

5 19. A polypeptide having the amino acid sequence of SEQ ID NO:16 analogues, homologues and derivatives thereof.

20 20. A polypeptide having the amino acid sequence of SEQ ID NO:17, analogues, homologues and derivatives thereof.

15 21. A polypeptide having the amino acid sequence of SEQ ID NO:18, analogues, homologues and derivatives thereof.

20 22. The polypeptide of any one of claims 1 to 21 and 100-101, wherein said polypeptide elicits an immune reaction that is specific to Streptococcal strains.

23. A polypeptide selected from the group consisting of:

(a) the HSP72 polypeptide having the amino acid sequence of SEQ ID NO:5;

25 (b) polypeptides that are immunologically reactive with antibodies generated by infection of a mammalian host with *Streptococcus pneumoniae* cells, which antibodies are immunologically reactive with the HSP72 polypeptide of paragraph (a);

30 (c) polypeptides that are capable of eliciting antibodies that are immunologically reactive with the HSP72 polypeptide of paragraph (a);

(d) polypeptides that are immunologically reactive with antibodies elicited by immunization with the HSP72 polypeptide of paragraph (a); and

(e) fragments of any of the foregoing polypeptides, either alone or in combination with other polypeptides to form a fusion protein.

24. The polypeptide of claim 23, wherein the polypeptides of paragraph (b) are selected from the group consisting of polypeptides of the genera *Streptococcus* and *Enterococcus*.

25. The polypeptide of claim 23, wherein the polypeptides of paragraph (b) are selected from the group consisting of polypeptides of the species *Streptococcus pyogenes*, *Streptococcus mutans*, *Streptococcus sanguis*, and *Enterococcus faecalis*.

26. The polypeptide of claim 23, wherein the fragments of paragraph (e) are selected from the group consisting of amino acids 439-607 of SEQ ID NO:5 (C-169); amino acids 527-541 of SEQ ID NO:5, and amino acids 586-600 of SEQ ID NO:5.

27. The polypeptide of claim 23, wherein the fusion protein of paragraph (e) is the Fucose Isomerase-HSP72 (C-169) protein having the amino acid sequence of SEQ ID NO:3.

28. A DNA sequence selected from the group consisting of:

- (a) the HSP72 DNA sequence of SEQ ID NO:4;
- (b) the HSP70 (DnaK) DNA sequence of SEQ ID NO:19;

(c) the HSP70 (DnaK) DNA sequence of SEQ ID NO:21;

(d) DNA sequences encoding polypeptides that are immunologically reactive with antibodies generated by infection of a mammalian host with *Streptococcus pneumoniae* cells, which antibodies are immunologically reactive with the HSP72 polypeptide (SEQ ID NO:5);

(e) DNA sequences encoding polypeptides that are capable of eliciting antibodies that are immunologically reactive with the HSP72 polypeptide (SEQ ID NO:5);

(f) DNA sequences encoding polypeptides that are immunologically reactive with antibodies elicited by immunization with the HSP72 polypeptide (SEQ ID NO:5);

(g) DNA sequences that are degenerate to any of the foregoing DNA sequences; and

(h) fragments of any of the foregoing DNA sequences, either alone or in combination with other DNA sequences to form a fusion DNA sequence.

29. The DNA sequence of claim 28, wherein the DNA sequences of paragraph (d) are selected from the group consisting of DNA sequences of the genera *Streptococcus* and *Enterococcus*.

30. The DNA sequence of claim 28, wherein the DNA sequences of paragraph (d) are selected from the group consisting of DNA sequences of the species *Streptococcus pneumoniae*, *Streptococcus agalactiae*, *Streptococcus pyogenes*, *Streptococcus mutans*, *Streptococcus sanguis*, and *Enterococcus faecalis*.

31. The DNA sequence of claim 28, wherein the DNA sequences of paragraph (d) are selected from the group

consisting of DNA sequences of the species *Streptococcus pneumoniae*, *Streptococcus agalactiae*, and *Streptococcus pyogenes*.

5 32. A DNA sequence of the formula of SEQ ID
NO:4 from nucleotide 682 to nucleotide 2502, or
derivatives thereof, coding for HSP72.

10 33. A DNA sequence of the formula of SEQ ID
NO:4 from nucleotide 1996 to nucleotide 2502, or
derivatives thereof, coding for the C-169 fragment of
HSP72.

15 34. A DNA sequence of the formula of SEQ ID
NO:4 from nucleotide 2050 to nucleotide 2502, or
derivatives thereof, coding for the C-151 fragment of
HSP72.

20 35. A DNA sequence of the formula of SEQ ID
NO:4 from nucleotide 2260 to nucleotide 2304, or
derivatives thereof.

25 36. A DNA sequence of the formula of SEQ ID
NO:4 from nucleotide 2437 to nucleotide 2481, or
derivatives thereof.

30 37. A DNA sequence of the formula of SEQ ID
NO:19 from nucleotide 204 to nucleotide 2027, or
derivatives thereof, coding for HSP70 of *Streptococcus*
agalactiae.

38. A DNA sequence of the formula of SEQ ID
NO:21 from nucleotide 248 to nucleotide 2074, or

derivatives thereof, coding for HSP70 of *Streptococcus pyogenes*.

39. A DNA sequence of the formula of SEQ ID
5 NO:25 from nucleotide 4 to nucleotide 456, or derivatives
thereof, coding for the C-terminal 151-residue fragment
(C-151) of HSP72.

40. A DNA sequence coding for a polypeptide
10 according to any one of claims 1-21 and 100-101.

41. A DNA sequence selected from the group
consisting of:

- (a) the HSP72 DNA sequence of SEQ ID NO:4;
- 15 (b) DNA sequences encoding polypeptides
that are immunologically reactive with antibodies
generated by infection of a mammalian host with
Streptococcus pneumoniae cells, which antibodies are
immunologically reactive with the HSP72 polypeptide (SEQ
20 ID NO:5);
- (c) DNA sequences encoding polypeptides
that are capable of eliciting antibodies that are
immunologically reactive with the HSP72 polypeptide (SEQ
ID NO:5);
- 25 (d) DNA sequences encoding polypeptides
that are immunologically reactive with antibodies elicited
by immunization with the HSP72 polypeptide (SEQ ID NO:5);
- (e) DNA sequences that are degenerate to
any of the foregoing DNA sequences; and
- 30 (f) fragments of any of the foregoing DNA
sequences, either alone or in combination with other DNA
sequences to form a fusion DNA sequence.

42. The DNA sequence of claim 41, wherein the
35 DNA sequences of paragraph (b) are selected from the group

consisting of DNA sequences of the genera *Streptococcus* and *Enterococcus*.

43. The DNA sequence of claim 41, wherein the
5 DNA sequences of paragraph (b) are selected from the group consisting of DNA sequences of the species *Streptococcus pyogenes*, *Streptococcus mutans*, *Streptococcus sanguis*, and *Enterococcus faecalis*.

10 44. The DNA sequence of claim 41, wherein the fragments of paragraph (f) are selected from the group consisting of nucleotide 1996-2502 (amino acids 439-607) of SEQ ID NO:4 (C-169); nucleotide 2260-2304 (amino acids 527-541) of SEQ ID NO:4; and nucleotide 2437-2481 (amino
15 acids 586-600) of SEQ ID NO:4.

45. The DNA sequence of claim 41, wherein the fusion DNA sequence of paragraph (f) is the Fucose Isomerase-HSP72 (C-169) DNA sequence of SEQ ID NO:1
20 (nucleotides 771-2912).

46. An expression vector including at least one DNA sequence according to claim 41 operably linked to a promoter.

25

47. A recombinant DNA molecule comprising a DNA sequence according to any one of claims 28 to 40, and one or more expression control sequence operably linked to the DNA sequence.

30

48. The recombinant DNA molecule of claim 47, wherein said expression control sequence is an inducible expression vector.

49. The recombinant molecule of claim 48,
wherein said expression vector comprises the λ PL
promoter.

5 50. A recombinant molecule according to claim
47 consisting of a plasmid selected from the group
consisting of: pURV3, pURV4, pURV5, pURV6, pJBD291,
pJBDA4, pJBdk51, pJBD177, pJBD171, pJBD177, pJBD179,
pJBDA1, pJBdf51, and pJBdf62.

10

51. A unicellular host transformed with an
expression vector of claim 46.

52. A unicellular host transformed with a
15 recombinant DNA molecule of claim 47.

53. A unicellular host according to claim 52,
wherein said host is selected from the group consisting
of: *E.coli* strains XLI Blue MRF', W3110, JM109, Y1090 and
20 BL21(DE3).

54. A method for producing a polypeptide or
fragment thereof comprising the steps of culturing the
unicellular host of claim 51 and isolating said
25 polypeptide or fragment.

55. An antibody or fragment thereof that
specifically binds to a polypeptide of claim 23.

30 56. An antibody or fragment thereof that
specifically binds to the epitope recognized by monoclonal
antibody F1-Pn3.1.

57. The antibody or fragment of claim 55, which is a monoclonal antibody.

58. The monoclonal antibody or fragment of claim 57, which is of murine origin.

59. The monoclonal antibody or fragment of claim 58, which is of IgG type.

60. The monoclonal antibody of claim 59, which is selected from the group consisting of F1-Pn3.1, F2-Pn3.2, F2-Pn3.3, and F2-Pn3.4.

61. The monoclonal antibody F1-Pn3.1.

62. A method for isolating the antibody of claim 55 comprising:

(a) introducing a preparation of the polypeptide of claim 23 into a mammal; and

(b) isolating serum from the mammal containing said antibody.

63. A method for isolating the monoclonal antibody of claim 57 comprising:

(a) introducing a preparation of the polypeptide of claim 23 to antibody producing cells of a mammal;

(b) fusing the antibody producing cells with myeloma cells to form hybridoma cells, and

(c) isolating said monoclonal antibody from the hybridoma cells.

64. A pharmaceutical composition comprising a polypeptide of claim 23.

65. The pharmaceutical composition of claim 64, which is a vaccine.

5 66. The pharmaceutical composition of claim 64, further comprising one or more pharmaceutically acceptable excipients.

67. A method for preventing infection of a
10 patient by *Streptococcus pneumoniae* or related bacteria comprising the administration of a pharmaceutically effective amount of the vaccine of claim 65.

68. A pharmaceutical composition comprising one
15 or more antibodies or fragments thereof according to claim 55.

69. The pharmaceutical composition of claim 68, which is a vaccine.
20

70. The pharmaceutical composition of claim 69, further comprising a pharmaceutically acceptable excipient.

25 71. The pharmaceutical composition of claim 69, wherein the antibody is selected from the group consisting of F1-Pn3.1, F2-Pn3.2, F2-Pn3.3, and F2-Pn3.4.

72. The pharmaceutical composition of claim 69,
30 wherein the antibody is F1-Pn3.1.

73. A method for treating a patient infected with or suspected of being infected with *Streptococcus pneumoniae* or related bacteria comprising the

administration of a pharmaceutically effective amount of the vaccine of claim 69.

74. A method for the detection of *Streptococcus pneumoniae* or related bacteria in a biological sample comprising:

- (a) isolating the biological sample from a patient;
- (b) incubating the antibody or fragment of claim 55 with the biological sample to form a mixture; and
- (c) detecting specifically bound antibody or fragment in the mixture which indicates the presence of *Streptococcus pneumoniae* or related bacteria.

75. The method of claim 74, wherein the antibody is selected from the group consisting of F1-Pn3.1, F2-Pn3.2, F2-Pn3.3, and F2-Pn3.4.

76. The method of claim 74, wherein the antibody is F1-Pn3.1.

77. A method for the detection of antibodies specific to *Streptococcus pneumoniae* or related bacteria in a biological sample comprising:

- (a) isolating the biological sample from a patient;
- (b) incubating a polypeptide of claim 23 with the biological sample to form a mixture; and
- (c) detecting specifically bound polypeptide in the mixture, which indicates the presence of antibodies specific to *Streptococcus pneumoniae* or related bacteria.

78. A method for the detection of *Streptococcus pneumoniae* or related bacteria in a biological sample comprising:

- (a) isolating the biological sample from a
5 patient;
- (b) incubating a DNA probe having the DNA sequence of claim 41 with the biological sample to form a mixture; and
- (c) detecting specifically bound DNA probe
10 in the mixture which indicates the presence of *Streptococcus pneumoniae* and related bacteria.

79. The method of claim 78, wherein the DNA probe is an oligomer having a sequence complementary to at
15 least about 6 contiguous nucleotides of a DNA sequence of claim 41.

80. The method of claim 79, which further comprises:

- (a) providing a set of oligomers which are
20 primers for a polymerase chain reaction method and which flank the target region; and
- (b) amplifying the target region via the polymerase chain reaction method.

25

81. The use of a pharmaceutically effective amount of the polypeptide of claim 23 for the prevention of *Streptococcus pneumoniae* or related bacterial infections in humans.

30

82. The use of a pharmaceutically effective amount of an antibody specific to HSP72 for the prevention of *Streptococcus pneumoniae* or related bacterial infections in humans.

83. A method for producing a polypeptide or fragment thereof comprising the steps of culturing the unicellular host of claim 52 or 53 and isolating said polypeptide or fragment.

84. A polypeptide in substantially pure form as obtained by the method of claim 83.

85. An antibody or fragment thereof that specifically binds to a polypeptide of claim 1 or 22.

86. A method for isolating the antibody of claim 86 comprising:

(a) introducing a preparation of the polypeptide of claim 1 or 22 into a mammal; and

(b) isolating serum from the mammal containing said antibody.

87. A pharmaceutical composition comprising a polypeptide of claim 1 or 22.

88. The pharmaceutical composition of claim 87, which is a vaccine.

89. The pharmaceutical composition of claim 87, further comprising one or more pharmaceutically acceptable excipients.

90. A method for preventing infection of a patient by *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae* comprising the administration of a pharmaceutically effective amount of the vaccine of claim 88.

91. An antibody or fragment thereof that specifically binds to a polypeptide of claim 1 or 22.

5 92. A method for the detection of *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae* in a biological sample comprising:

(a) isolating the biological sample from a patient;

10 (b) incubating the antibody or fragment of claim 91 with the biological sample to form a mixture; and

(c) detecting specifically bound antibody or fragment in the mixture which indicates the presence of *Streptococcus pneumoniae*, *Streptococcus pyogenes* or
15 *Streptococcus agalactiae*.

93. A method for the detection of antibodies specific to *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae* in a biological sample comprising:

20 (a) isolating the biological sample from a patient;

(b) incubating a polypeptide of claim 1 or 22 with the biological sample to form a mixture; and

(c) detecting specifically bound
25 polypeptide in the mixture, which indicates the presence of antibodies specific to *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae*.

94. A method for the detection of *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae* in a biological sample comprising:

30 (a) isolating the biological sample from a patient;

(b) incubating a DNA probe having the DNA sequence of claim 28 with the biological sample to form a mixture; and

(c) detecting specifically bound DNA probe
5 in the mixture which indicates the presence of *Streptococcus pneumoniae*, *Streptococcus pyogenes* or *Streptococcus agalactiae*.

95. The method of claim 94, wherein the DNA
10 probe is an oligomer having a sequence complementary to at least about 6 contiguous nucleotides of a DNA sequence of claim 28.

96. The method of claim 95, which further
15 comprises:

(a) providing a set of oligomers which are primers for a polymerase chain reaction method and which flank the target region; and

(b) amplifying the target region via the
20 polymerase chain reaction method.

97. The use of a pharmaceutically effective amount of the polypeptide of claim 1 or 22 for the prevention of *Streptococcus pneumoniae*, *Streptococcus*
25 *pyogenes* or *Streptococcus agalactiae* infection in humans.

98. The use of a pharmaceutically effective amount of an antibody specific to HSP72 for the prevention of *Streptococcus pneumoniae*, *Streptococcus pyogenes* or
30 *Streptococcus agalactiae* infection in humans.

99. The use of a pharmaceutically effective amount of a polypeptide according to any one of claims 2

to 21 for the prevention of Streptococcal infections in humans.

100. A polypeptide having the amino acid
5 sequence of SEQ ID NO:23, analogues, homologues, or
derivatives thereof.

101. A polypeptide having the amino acid
sequence of SEQ ID NO:24, analogues, homologues or
10 derivatives thereof.

1/22

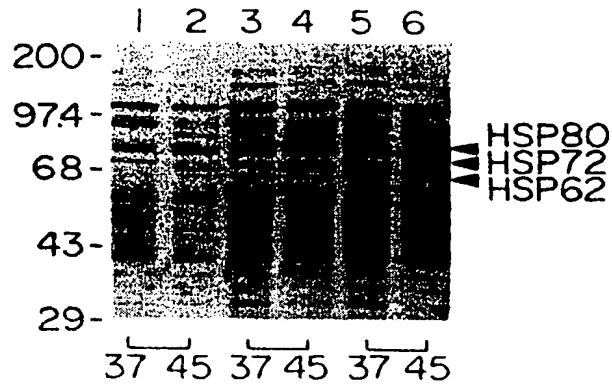


Fig. 1A

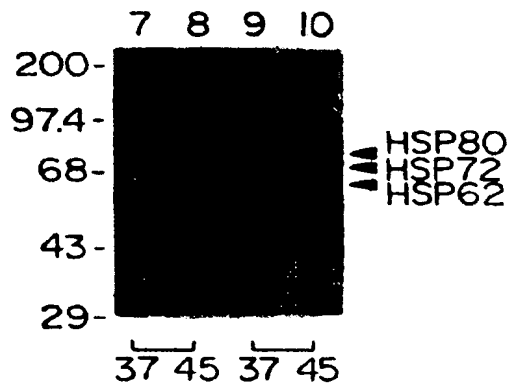


Fig. 1B

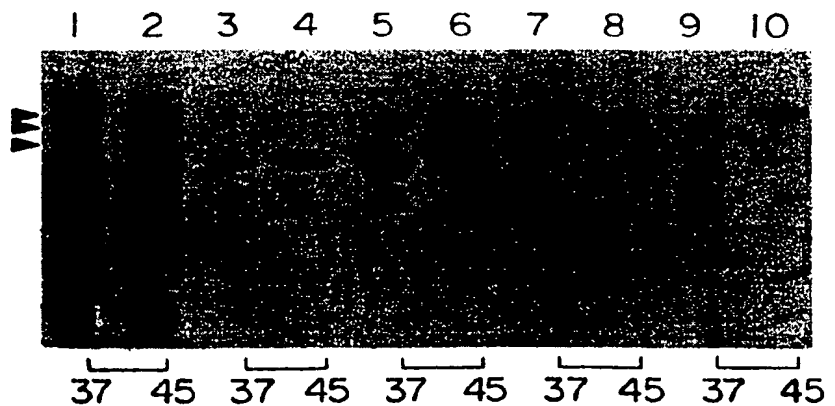
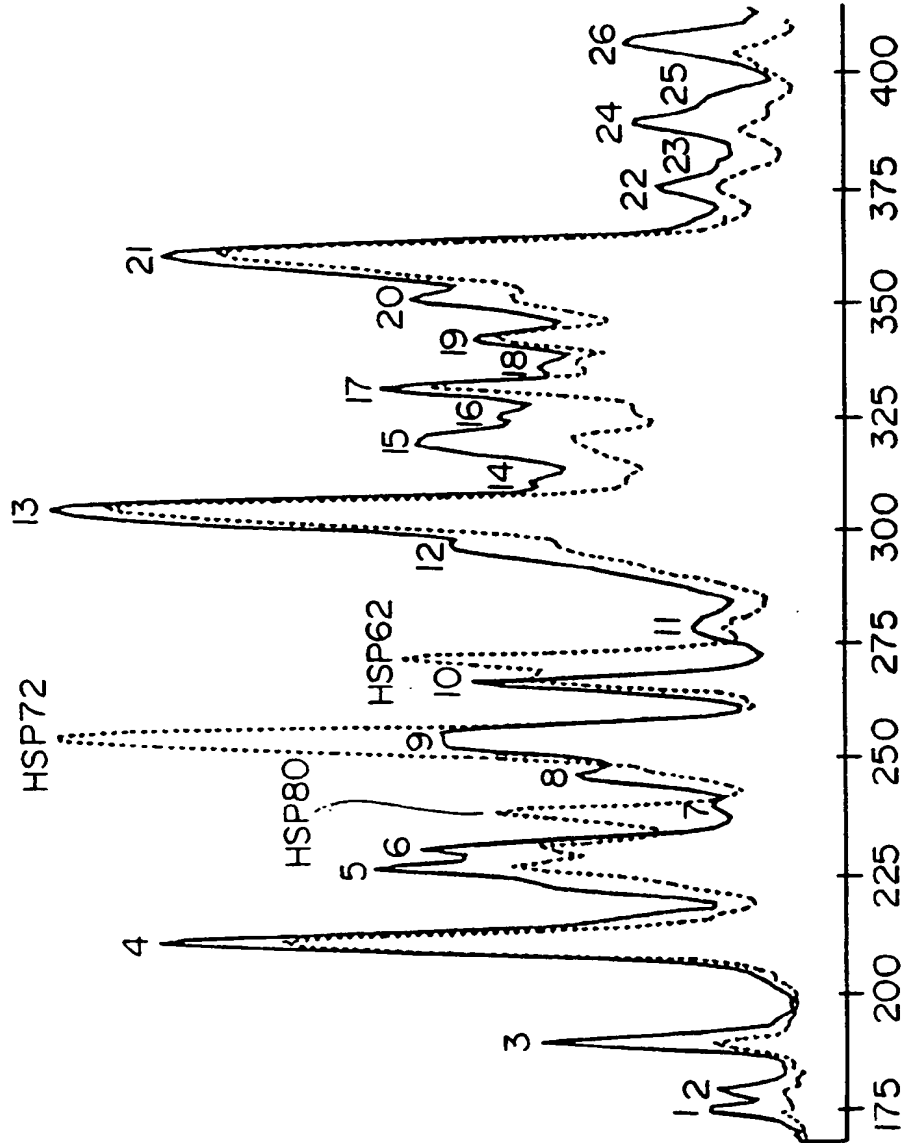


Fig. 3

2/22



DISTANCE (mm X 10) FILE: 2

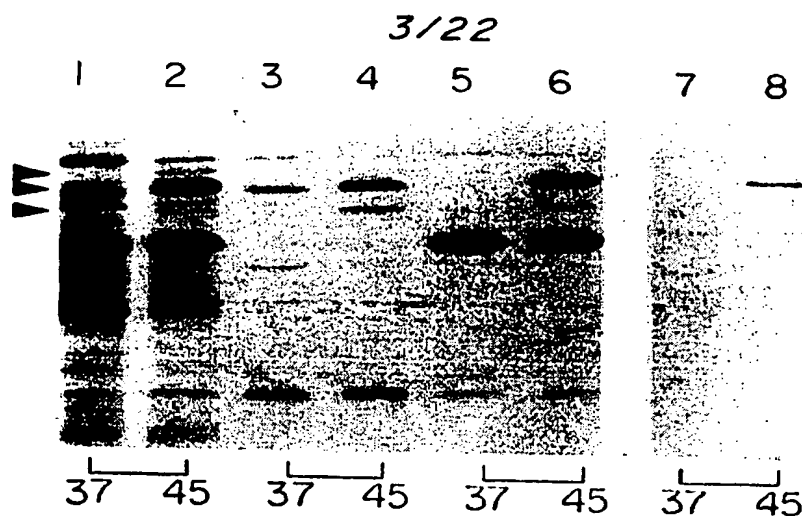


FIG. 4

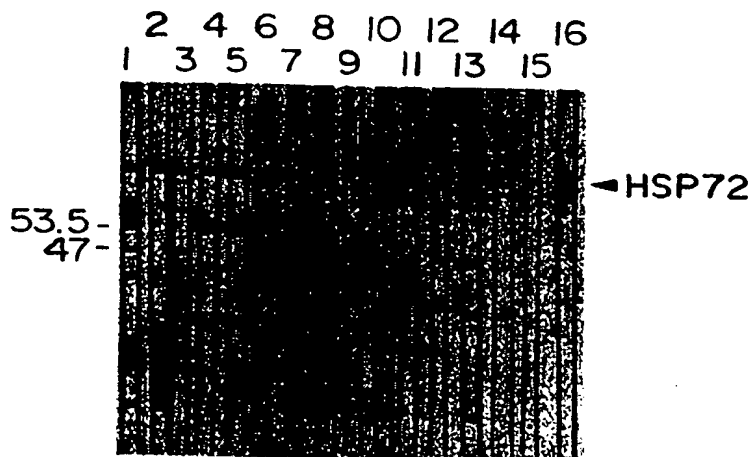


FIG. 5A

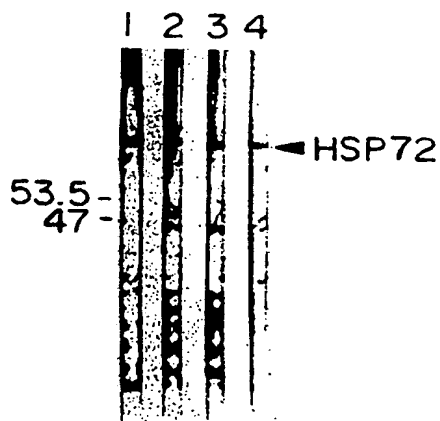
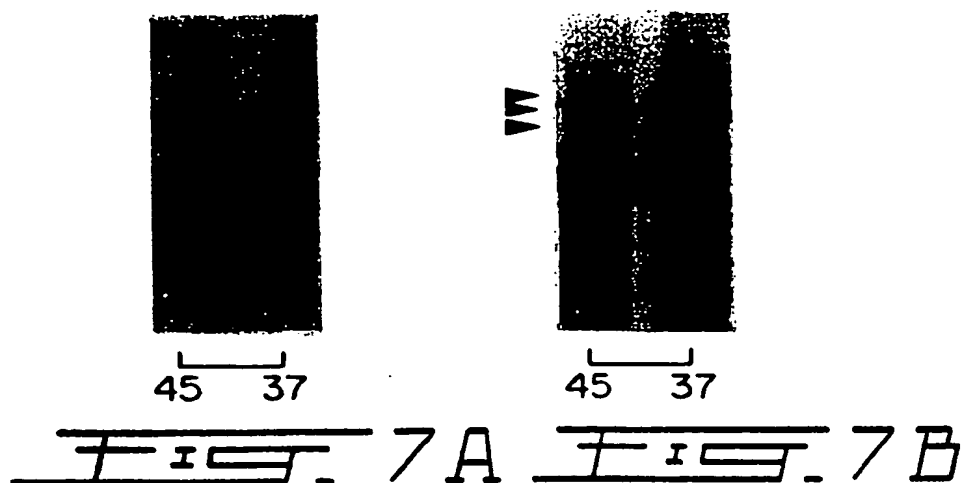
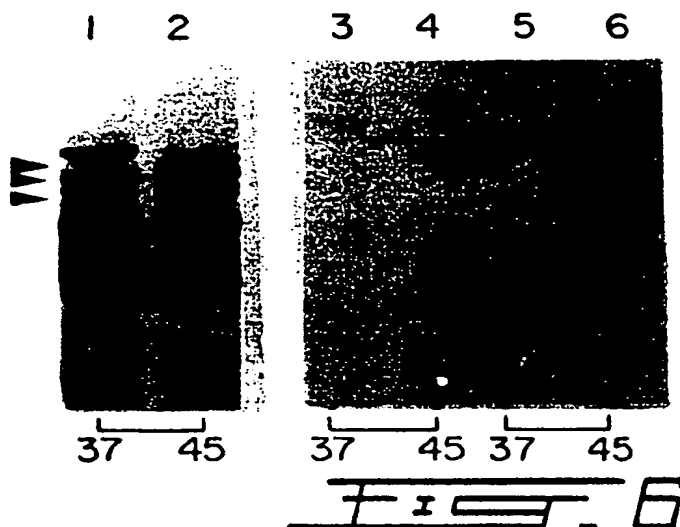


FIG. 5B

4/22



SUBSTITUTE SHEET (RULE 26)

5/22



FIG. 8

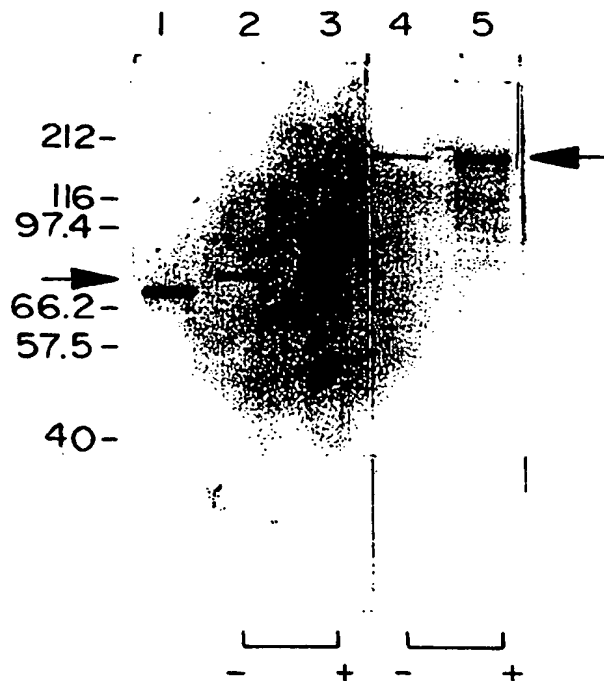
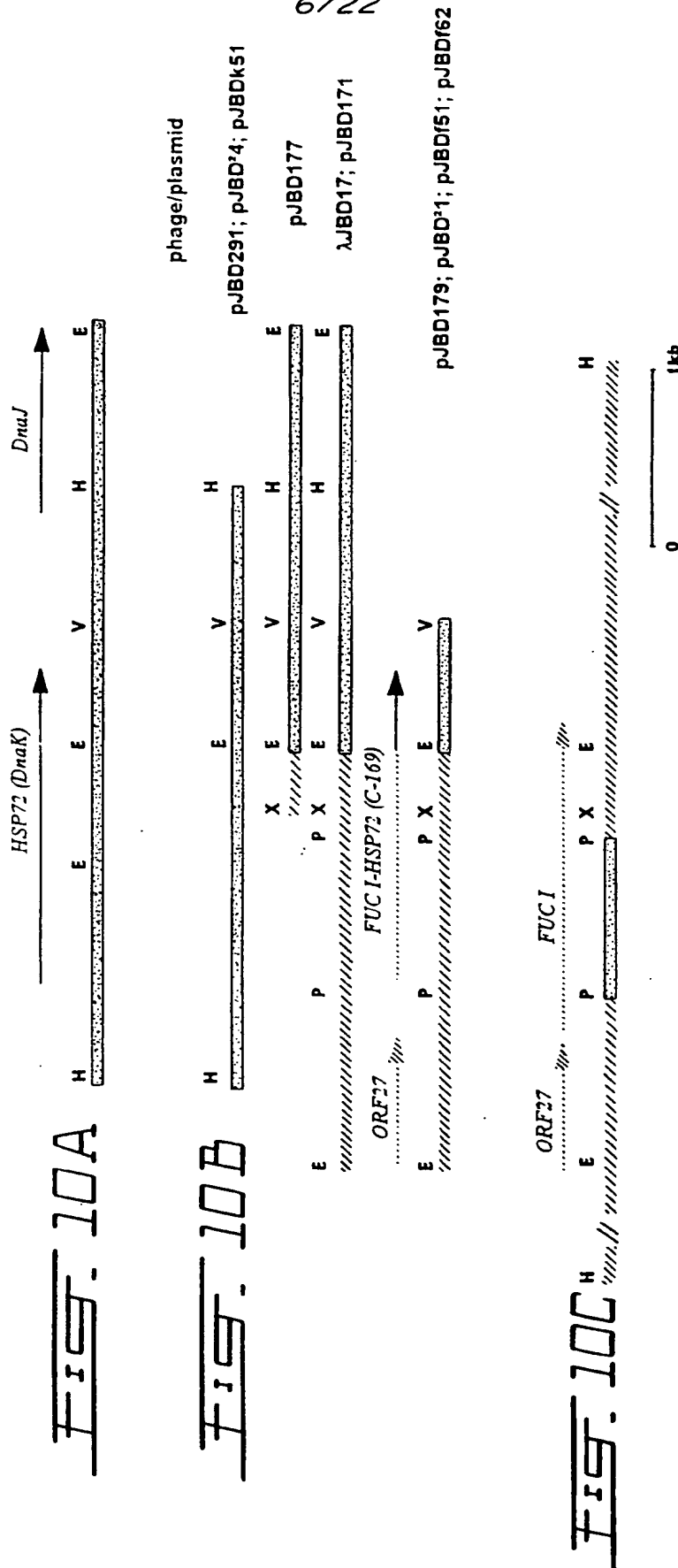


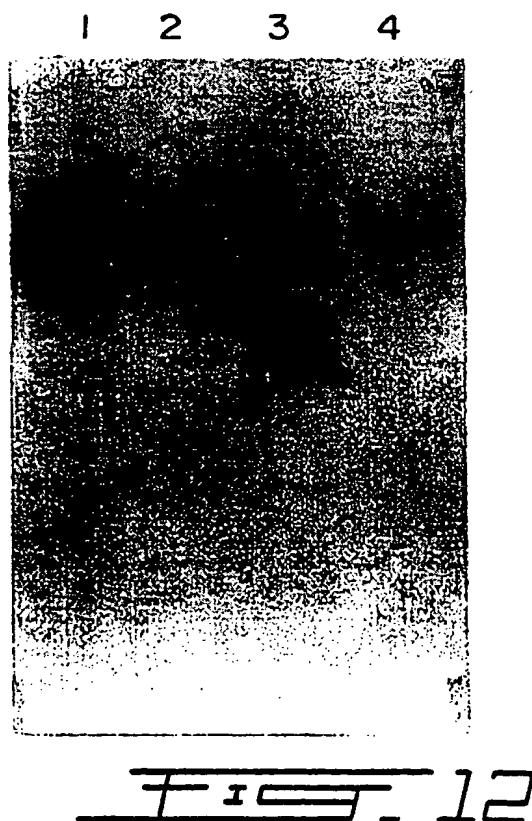
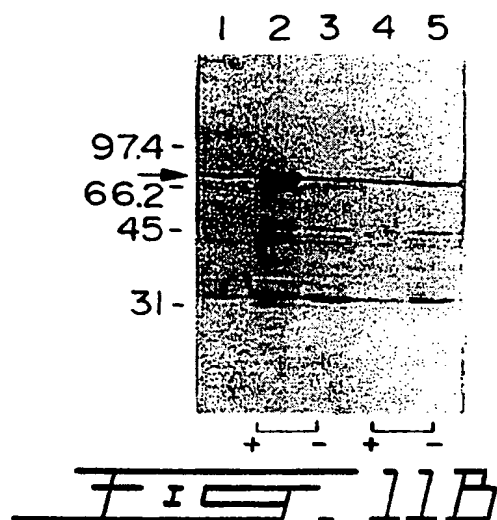
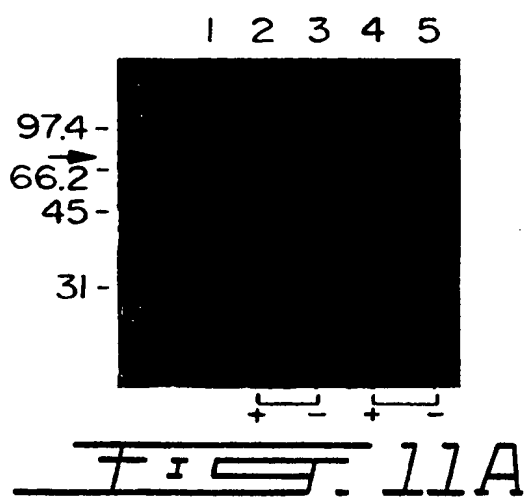
FIG. 9

SUBSTITUTE SHEET (RULE 26)

6/22



7/22



SUBSTITUTE SHEET (RULE 26)

8/22

Consensus	M-----SKI	IGIDLGTINS	QAVLEG..P	KVI.N.EG	R	TTPS	..VAF-K	50
DNAK ECOLIG..IMD.TT.	R.LE.A..D.	II.YTO	43	
DNAK BOREUG..IM.HGK.	V..Q.S..G.	I..YIN	44	
DNAK BRUOVA.VMD.KNA	..E.A..A.	II..TD	44	
DNAK CHLPN	..SEHKKS..S.M..GOA	..TSS..I.	I.....	49	
DNAK BACMEGE..	..P.P..N.	V.....	43	
DNA2 BACSUVGE..	..A.A..N.	V.....	43	
DNAK STAAUT..DE.	..Q.P..S.	V.....	43	
HSP72 SPNEU	A.....TES	..I.A.P..N.	V.S...	43	
DNAK LACLA	A.....TES	..I.P.P..N.	V.....	43	
DNAK MYCTUARA V	V.S....GD.	V.VA.S..S.	I...AR	44	

Consensus	NCE..VG..A	KRCVAVNP..	T..SIR..MG	-----	-----	..KV-	100
DNAK ECOLI	D..TL..OF.QN	..LFA..LI.	RRFQDEEVOR	DVSIMPF.II	93	
DNAK BOREU	K..RL..OV.	..N.M..EN	..IY...F..	RRFEE..VAS	EIKMVPY.IE	92	
DNAK BRUOV	GD..GVA..OF.EG	..LFAV..LI.	RRYDDPMVIK	DKDLVPY.IV	94	
DNAK CHLPN	GN..KL..IH.EK	..LG.T..FT.	RKYSE..VAS	EIQTVPYT.T	97	
DNAK BACME	..RO..EV.	..I..N-	..II.V..H.	TDH...	77	
DNA2 BACSU	..RO..EV.	..SI..N-	..IM...H.	TDY...	77	
DNAK STAAU	..TO..EV.	..I..N-	..VO...H.	TDY...	77	
HSP72 SPNEU	..II..DA.D-	..VI..SK.	TSE...	77	
DNAK LACLA	..II..DA.E-	..II..SK.	TSE...	77	
DNAK MYCTU	..VL..OF.	..N...VDR	..MR.V..H.	SDWSI..	79	

Consensus	-----E	..G--K.YTP	..ELSA..IQ.	LR.TAE	YLG E.VT	AVITV	150
DNAK ECOLI	AADNGDAWV.	VK...QKMA.	PQ...EV	KK M.K..D.	..P..E....	141	
DNAK BOREU	KGLNGDARVN	ISNIK.QMS.	P...AT	TK M.E..A.	..K..E....	142	
DNAK BRUOV	KGLNGDAWV.	VH...K.S.	SQ...M..K	M.E..S.	..T..C....	142	
DNAK CHLPN	SGSKGDVAF.	VD...Q..E.	G.Q..MK	M.E..A.	..T..E....	145	
DNAK BACME	AE...Q..Q.M.	I...H..GY.	E...P..K....	116		
DNA2 BACSU	IE...D..Q.V.	I...H..SY.	S...T..SK....	116		
DNAK STAAU	D IE...S..Q..M.	N..N..G	..K..DK....	116		
HSP72 SPNEU	S AN...E...Q..M.	Y..GY..D	..K..K....	116		
DNAK LACLA	S AN...E...Q..M.	N..A..G	..K..EK....	116		
DNAK MYCTU	ID...K..A	P...R..MK	..RD..A..DI	D....T	118	

Consensus	PAYFNDACRO	ATKDACKIAG	LEV..RIINEP	TAAALAYGLD	R.....ILV	200
DNAK ECOLIR..K..	GIGNRT.A	191
DNAK BOREUK..V.IE	KHEE-IVA	191
DNAK BRUOVL..	SEK-T.A	191
DNAK CHLPNS..A	S.....R..D.K..P.I.	VCDK-K.A	194
DNAK BACMEE..E..E	IDEDQTV..	166
DNA2 BACSUE..E..	IDEDQT..	166
DNAK STAAUE..E..	IDKDEKV..	166
HSP72 SPNEUE..V.	IDKEEK..	166
DNAK LACLAE..V.	IDKDEK..	166
DNAK MYCTUC..N..L..V.PG..	GEKEOR..	168

Fig. 13A

9/22

Consensus	.DLGGGTFLV	SILELGD--G	--VFEV	ST.	GDN	LGDDF	DQ.IID.LV.	250
DNAK ECOLI	Y.....I	.I.IDEVD	EKT...	LA.N	.TH...	E...	SRL.NY..E	241
DNAK BOREU	Y.....I	K.N	.TH...	.N.	DE..KH.IS	237
DNAK BRUOV	Y.....I	.V..I....	K.N	.TF...	E...	IRLVEY..A	237
DNAK CHLPN	F.....II....	L.N	.TL...	EV..KWMIE	240
DNAK BACME	Y.....	RA.A	.R....	V..Y..A	212
DNA2 BACSU	Y.....	R.A	.R....	V..H..S	212
DNAK STAAU	F.....	L.A	.K....	V..Y..A	212
HSP72 SPNEU	F.....	D.L.A	.K....	K..H..A	212
DNAK LACLA	F.....	D.L.A	.N.K...	K..WM.A	212
DNAK MYCTU	F.....	.L..I.E...	...V..RA	S	.H....	.W	RVV.W..D	214

Consensus	EFKKE.GIDL	S.DKVALRL	KDAAEKAK..	LSGV..T.I.	LPFITTA.-..	300
DNAK ECOLI	...DO...	RN.PL.M...	.E.....	IE	SAGQ.DVN	Y...D.AT
DNAK BOREU	...SA...	.N.....	.E.....	IE	ACEAS.N	...D.AN
DNAK BRUOV	...S...	KN..L....	IE	SSQO.E.N	...D.OT
DNAK CHLPN	...QE...	.K.N.....	IE	SS.E.N	Q...MD.AQ
DNAK BACME	...N.V...	.K.....	KD	TS.Q.S	...G.EA
DNA2 BACSU	...N...	.K.....	KD	SS.Q.S	...G.EA
DNAK STAAU	...N.V...	.Q.....	KD	SO.Q.S	...S.G.EN
HSP72 SPNEU	...N...	.T...M...	KD	TS.Q.S	...G.EA
DNAK LACLA	...N...	.Q.....	KD	TT.Q.S	...G.AA
DNAK MYCTU	K...GTS...	TK...M...	RE.....	IE	SSQS.S.N	...V.VDADK

Consensus	GFLHL...LT	RAKFE.L...	LV.RT..E...	.AL	DAGLS.	S.ID.	MLVG	350
DNAK ECOLI	..K.MNIKV.	..L.S.VED	.N.SIE.LK	V..Q....	V	.D..D....		340
DNAK BOREU	..K..QYT..	...QVNH	.OK.KE	CL	K.IK...	.KA	.D.NE....	336
DNAK BRUOV	..K..AIK.S	..NV.S.VDD	A.HA.VE	CK	A..K...	.KA	GE..E.M...	336
DNAK CHLPN	..K..ALT..	..Q..K.AAS	.IE..KS	CI	K..S...	.K..A	KD..D.L...	339
DNAK BACME	..K..EVS.S	..DE.SAG	.E..MA	VR	Q..K....	A	EL..K....	311
DNA2 BACSU	..K..ELT..	..E.SSH	.E..MG	VR	Q..Q....	A	E..K....	311
DNAK STAAU	..K..EVN..	..S..E.SDS	.IR..ME	IR	Q..MK...	TN	D..E....	311
HSP72 SPNEU	..K..EMT..	..DD.TRD	.E..KV	VR	Q..S....	L	E..E....	311
DNAK LACLA	..K..EVA..	..DE.THD	.EA.RQ	VR	Q..S....	T	D..E....	311
DNAK MYCTU	N..F.DEQ..	..E.QRITQD	.LD..RK	FQ	SVIA	T.I.V	E..H..M...	314

Consensus	GSTRIPAVQ.	.VK...G-KE	PNRGNPDEV	VA	GATCGG	VLIGDMKDW	400
DNAK ECOLI	..Q..M.M..K	K.AEFF....	.R.D....	A	..M....	..L	389
DNAK BOREU	..Q..I.K..I	.DIF..QD	A	..M....	I...ET..M.	385
DNAK BRUOV	..M..M.EI.K	C..AFF....	.H.....	..M....	..M....	Q...L	385
DNAK CHLPN	..MS.M....	E.T.ELF...M....	..M....	G.E...L	388
DNAK BACME	..D..AI.KET	.QD	.H.....	..L....	..L....	..L	360
DNA2 BACSU	..E..AI.KET	..AHL....	..L....	I...L	360
DNAK STAAU	..E..A..KEIM....	..M....	I...L	360
HSP72 SPNEU	..VE..A..AETS....	..M....	..M....	I...L	360
DNAK LACLA	..VE..L.RHEINS....	..M....	..M....	I...L	360
DNAK MYCTU	..M..TD	L..ELT.G..M....	..M....	K.E...L	364

Fr 13B

10/22

Consensus	LLDMIFLSLG	DET	GGV.T.	ILIERNTTIPT	.KSQVFSTAA	DNQ	AVDIHV	450
DNAK ECOLIM.	M.T	..AK	E	..S..T...	439
DNAK BOREUL.	M.K	K.TS...K.	435
DNAK BRUOVL.	F.R	K..T	E	..S..T.R.	435
DNAK CHLPNL.	M.T	V.	Q.K.I.	..P..T.V.	438
DNAK BACMEM.	F.K	S.S..T...	410
DNA2 BACSUM.	F.K	D.	S.	..T...	410
DNAK STAAUIL	FMNT	S..IY..	V	..PS..V..	410
HSP72 SPNEUM.	F.K	D.	S.	..P...	410
DNAK LACLAM.	F.K	D.	S.	..P...	410
DNAK MYCTUK.	M.R	KR.ET.	D	..PS..Q.Q.	414

Consensus	L	G	E	R	M	A	A	D	N	K	L	G	R	F	I	T	D	I	P	R	A	R	G	P	Q	I	E	V	I	F	D	I	L	N	G	I	V	M	A	K	D	500
DNAK ECOLI	K	R	.	.	.	S	.	.	Q	.	N	.	D	.	G	.	N	.	.	M	A	.	D	.	L	H	.	S	.	.	489
DNAK BOREU	E	.	.	Q	.	R	.	.	N	.	T	.	D	.	G	.	A	.	.	V	.	.	S	.	.	A	.	.	.	H	.	S	.	.	485		
DNAK BRUOV	F	.	A	S	V	K	L	.	.	.	L	.	.	Q	.	D	.	V	.	G	V	R	.	.	S	.	A	.	.	.	N	.	S	.	.	485		
DNAK CHLPN	K	.	.	.	E	I	S	.	.	A	.	.	F	H	.	S	.	.	488		
DNAK BACME	S	T	V	.	.	.	S	.	.	K	.	.	N	.	S	.	.	460			
DNA2 BACSU	S	T	V	.	.	.	S	.	.	K	.	.	N	.	S	.	.	460			
DNAK STAAU	T	E	.	K	.	.	.	S	.	K	.	.	N	.	S	.	.	460			
HSP72 SPNEU	T	A	.	I	.	.	.	S	.	K	.	.	S	.	S	.	.	460			
DNAK LACLA	T	A	.	I	.	.	.	S	.	K	.	.	S	.	S	.	.	460			
DNAK MYCTU	Y	E	I	.	H	.	L	.	S	.	E	.	.	G	A	H	.	I	.	.	464			

Consensus	.GT	KEQ	II	II	SSSG	ISD	ETIDRMVKA	AE	ANAEADKKR	EEV	..R	NEAD	550					
DNAK ECOLI	KNSG	..K	T	..KA	NED	..OK	R	R	FE	..L	QT	..QG	539			
DNAK BOREU	M	..G	..K	R	..E	ESH	E	..LK	..NTEAK	..T	N	535			
DNAK BRUOV	G	VP	..H	Q	R	..QA	G	..A	D	ER	R	..S	EAK	..Q	535		
DNAK CHLPN	V	ASG	..K	R	..EA	QED	..Q	..R	I	K	E	..R	ASDAK	538		
DNAK BACME	L	..N	..A	T	..K	..T	..DE	E	..D	..Q	K	..EL	510		
DNA2 BACSU	L	..G	..N	T	..K	E	..E	E	..D	..A	KK	..IEV	510	510		
DNAK STAAU	L	..N	..R	T	..Q	..S	..EV	R	..DL	510	510				
HSP72 SPNEU	L	..Q	..T	V	..Q	N	..T	E	M	S	..K	..DL	..V	510		
DNAK LACLA	L	..Q	..T	V	..K	N	..E	..K	MD	..A	..K	..DT	510	510		
DNAK MYCTU	K	..G	..N	T	..R	..QEG	KE	D	..I	H	..E	R	..R	ADV	..Q	514

Consensus	.LVF.TER	L	K....K..E.	..K..E.A..	.IK.ALE..D	...-IKAK..	600
DNAK ECOLI	H.IHS.R.OV	EEAGD.LPAD	DKTAL.S.LT	A.ET..KGE.	KAA..E..MQ	588	
DNAK BOREU	S.IYQ...S.	.EYSE.ISSE	DKEAL.SKIK	E..ES..KE.	-ISL..SRTE	584	
DNAK BRUOV	S..HS...S.	AEYGD.VSAD	DK.AI.D.IA	A..TS..GE.	-AED....TO	584	
DNAK CHLPN	S.MI.RA...AI	.DYKEQIP.T	LV.EI.ERIE	NVRN..KDDA	PIEK..EVTE	588	
DNAK BACME	Q...T...T.	.DLEG.VE.A	EVTKANE.KD	A..A.I.KN.	LEE....KD	559	
DNA2 BACSU	Q...Q...T.	.DLEG.VD.E	OV.KAND.KD	A..A.I.KNE	FEE....KD	559	
DNAK STAAU	S...QV...T.	TDLGENIG.E	DK.SA.EKKD	A..T...GQ.	IED...S.KE	559	
HSP72 SPNEU	QAI.A...TI	.ETEG.GFDA	ERDAAQA.LD	D..K.Q.DNN	LID.M...LE	559	
DNAK LACLA	A...Q...T.	.DLEG.VE.A	EV.KA.D.KE	E..K..GE.	IDD....SE	559	
DNAK MYCTU	T..YQ...FV	.EQRE--A.G	GS.VP.DTILN	KVDA.VAEAE	-GG.-TWRIG	559	

FIG. 13C

11/22

Consensus	.L....Q.L. ...-YE----	-A.AQ... ..-----	650
DNAK ECOLI	E.AQVS.K.M	ETA.-Q.... .QCHAO.QTA	GADASANNAK DDD..... 624
DNAK BORBU	E.QKASYKIA	EMM..KDSQ	QN.NSQ.ENG PQSNTSEEGK EADY..... 627
DNAK BRUOV	A.AEVSMK.G	QAM...AAQA	AEGAG.EGGE QASSSKDDV DADY..... 627
DNAK CHLEN	D.SKHM.KIG	ESMQSQSASA	AASSA.NAKG GPNINTEDLK KHSFSTKPPS 638
DNAK BACME	E.QEIV.A.T	VKL.....	.Q.QQ..-Q- ---A---GE OGA.....Q 588
DNA2 BACSU	E.QTIV.E.S	MKL.....	.E.AK..-Q- ---AQQGANA EGK.....A 592
DNAK STAAU	E.EKVI.E.S	AKV.....	.Q.AQ..-QQ AQQANAGQNN DST..... 595
HSP72 SPNEU	A.NEKA.G.A	VKL.....	.Q.AA.---- ---QAQEGA EGA.QATGNA 595
DNAK LACLA	A.SETA.N.A	VKL.....	.Q.NA..GEA SEATDAQEGP KDA..... 596
DNAK MYCTU	YFGHQVGDGE	AGFGVA....	.GSGASDLRS SSGCVIGHWR CPP..... 597
Consensus	---...D.E.. E...-----	..	672
DNAK ECOLI	..-W.A.FE	.VKD.....	KK 637
DNAK BORBU	..-----EVVD	..-----DKK	-- 635
DNAK BRUOV	..-----E.ID	D-----NKK	SS 637
DNAK CHLEN	NNGSSEDHIE	.ADVETIDND	DK 660
DNAK BACME	NDDV.A.FE	.VNDD.....	KK 605
DNA2 BACSU	DDNV.A.YE	.VNDDQN...	KK 611
DNAK STAAU	..-VE.A.FN	.VKDDDK...	-- 609
HSP72 SPNEU	GDDV.G.FT	.K--.....	-- 607
DNAK LACLA	..NIF.GDFE	.SK.....	-- 607
DNAK MYCTU	..RRRAGRCF	PRLG.....	-- 609

12/22

1 2 3

106-
80-
49.5-
32.5-
27.5-
18.5-

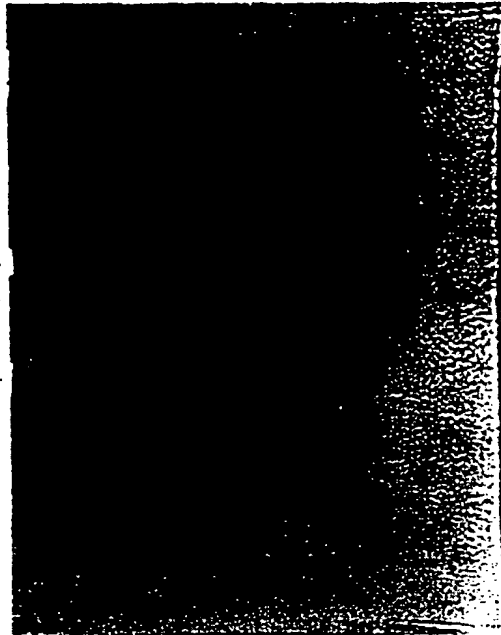


FIG. 14

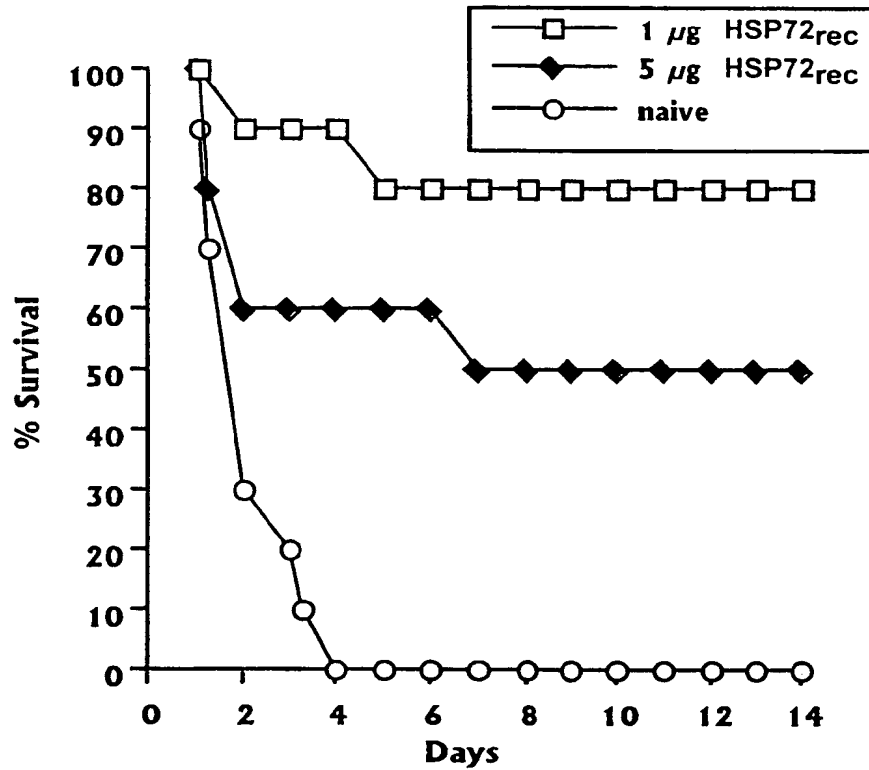
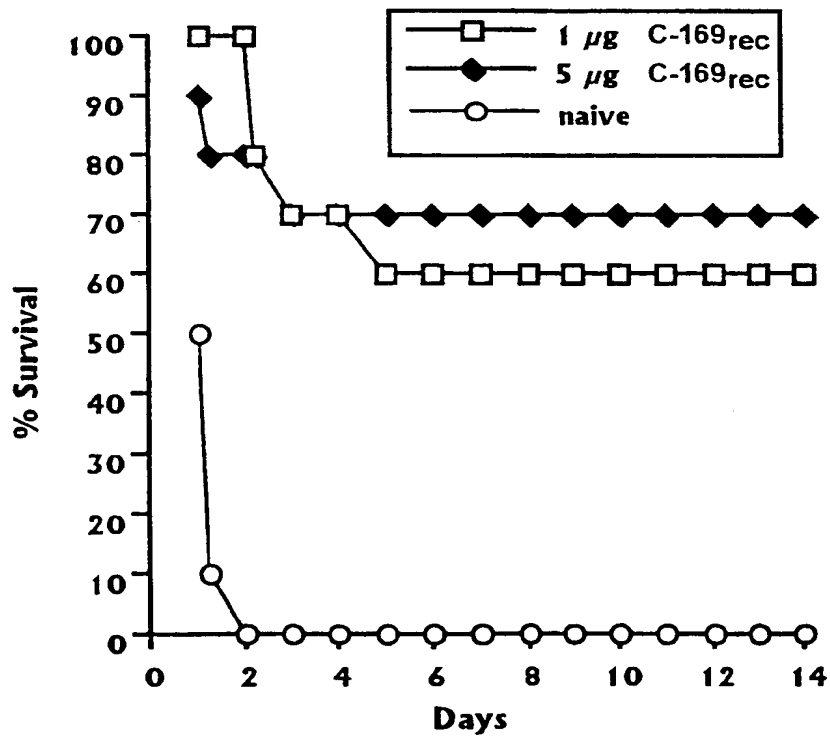
1 2 3 4 5



FIG. 15

SUBSTITUTE SHEET (RULE 26)

13/22

FIG. 16FIG. 17

14/22

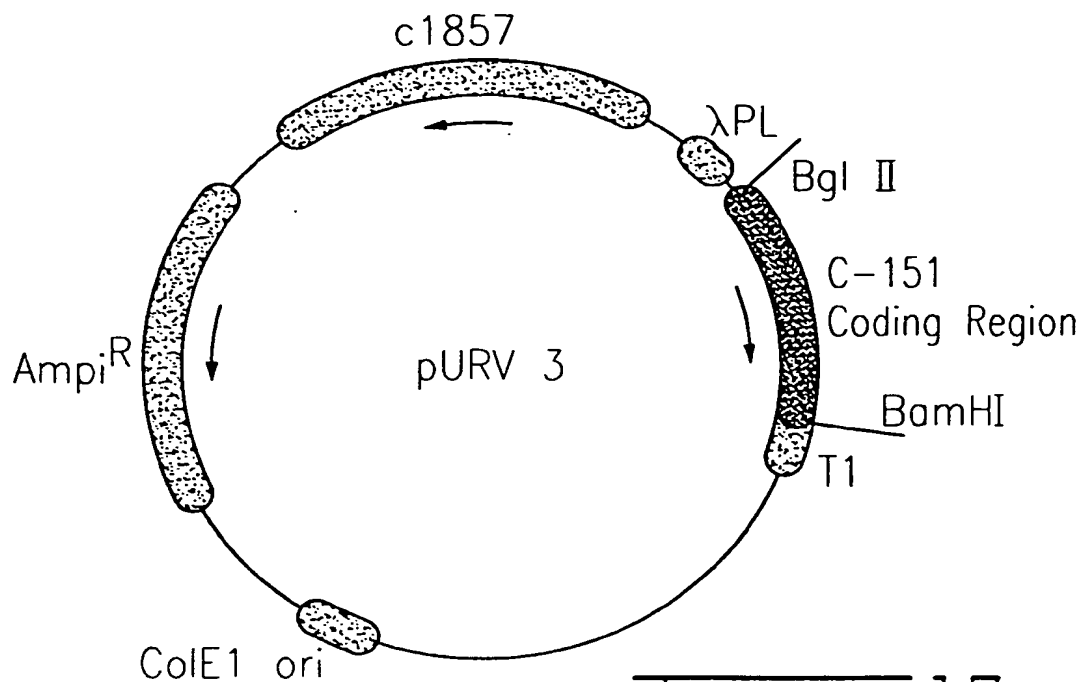


FIG. 18

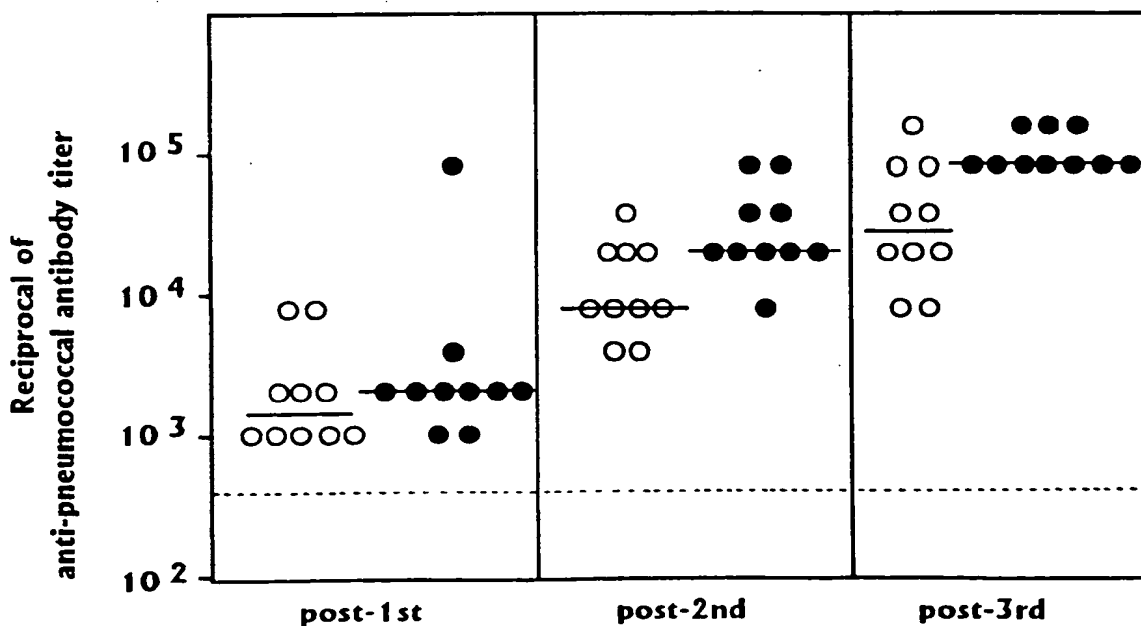


FIG. 19

15/22

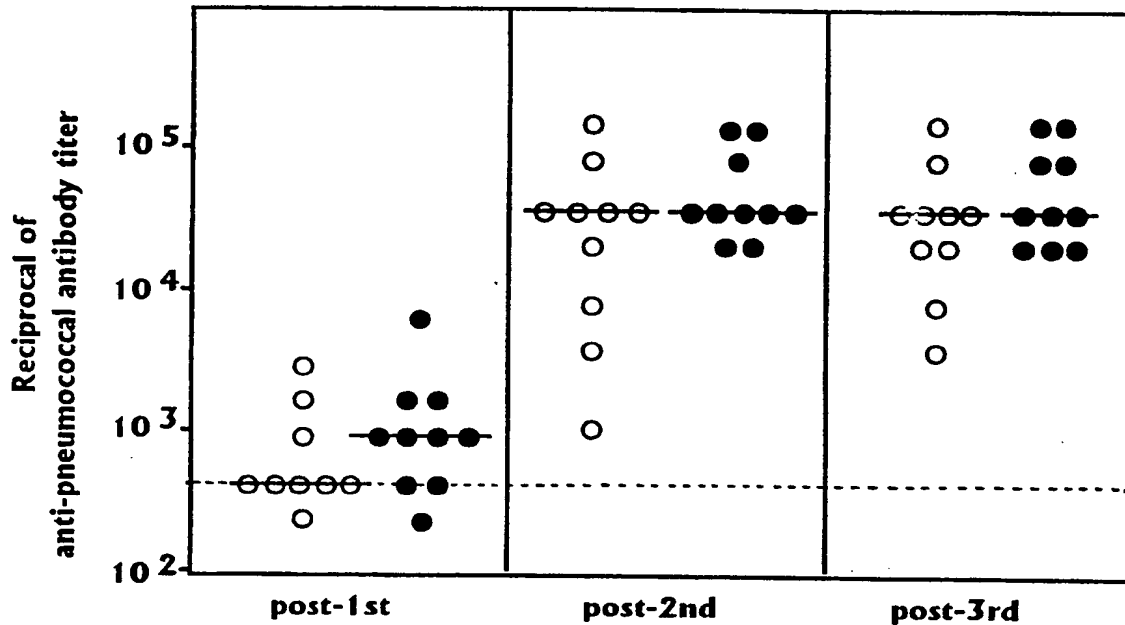


Fig. 20

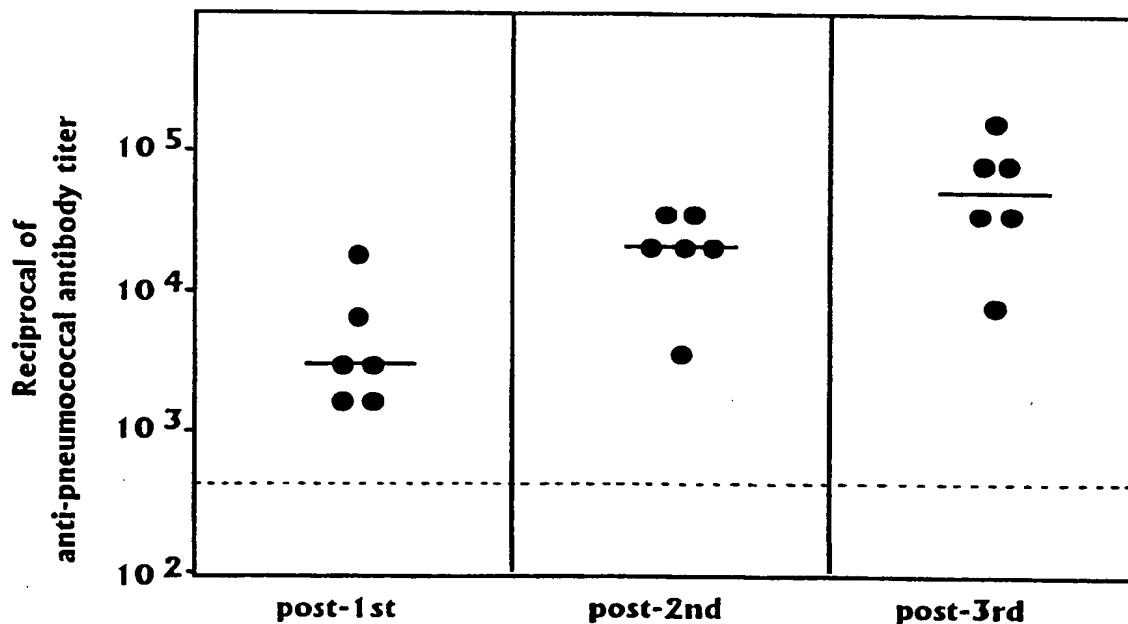


Fig. 21

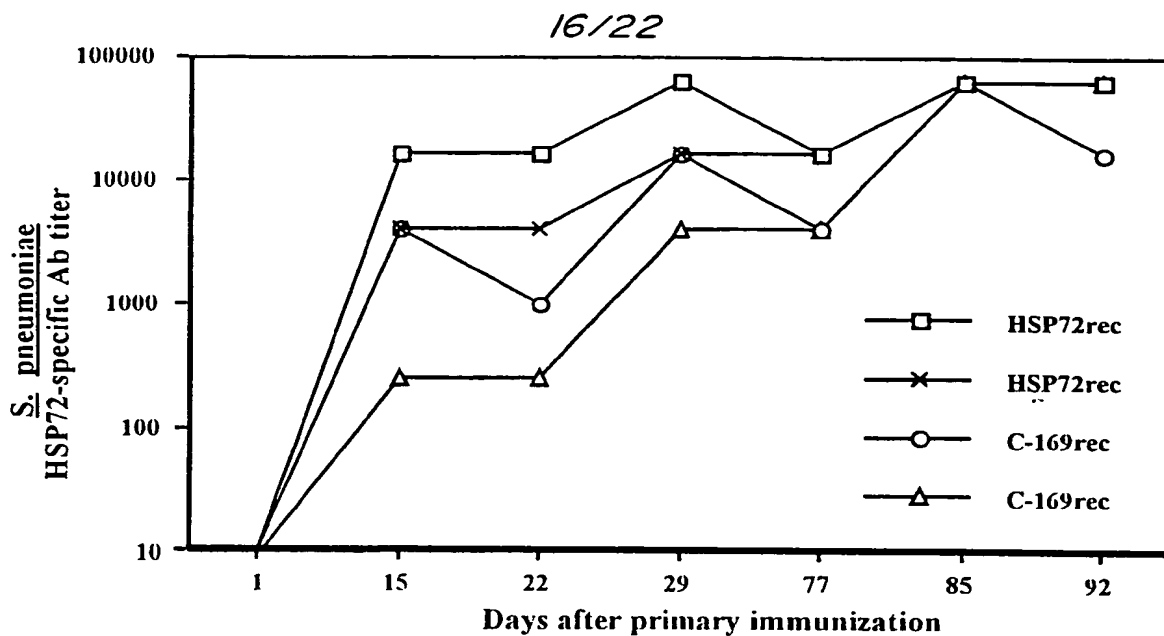


FIG. 22

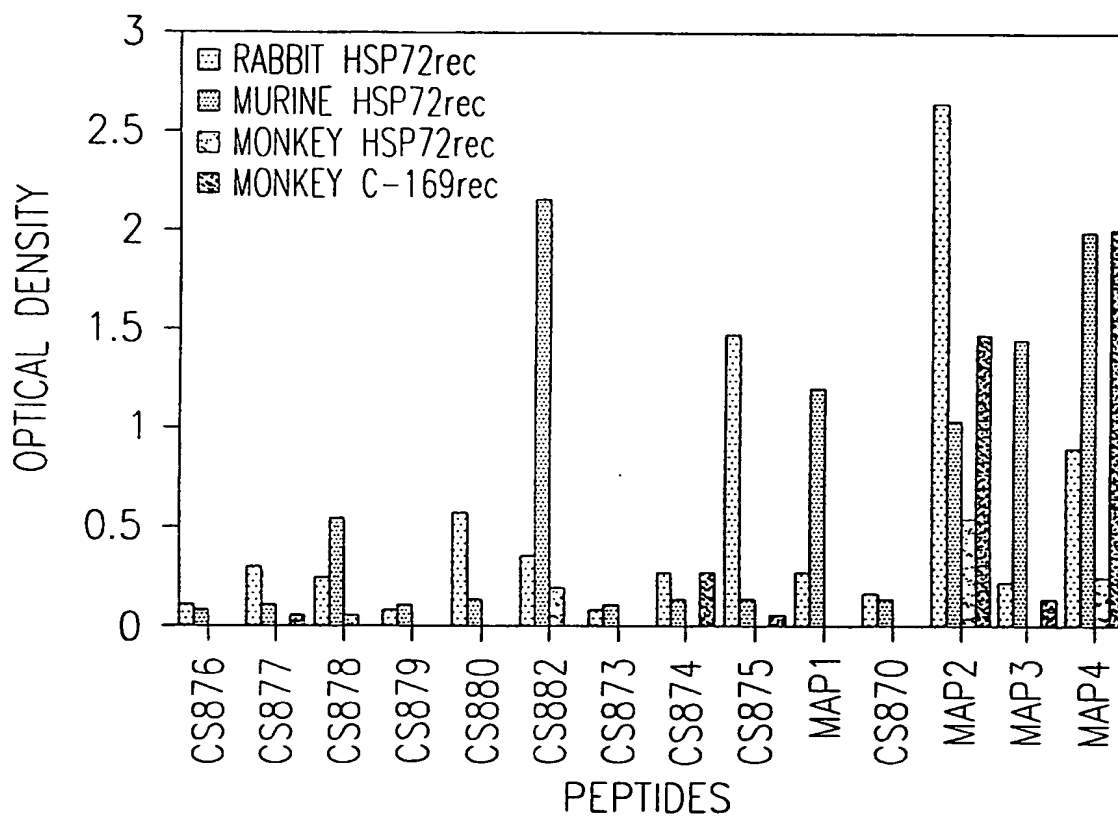


FIG. 23

17/22

spn-orfC.....T..	50
sga-orfT.....A..	50
sgb-orfT.....A..	50
Consensus	ATGTCTAAAA TTATYGGTAT TGACTTAGGT ACAACAAACT CAGCACTWGC	50
spn-orfA.....A GC.....C..A.....A..C.	100
sga-orfG.....T CA.....T..T.....C..T.	100
sgb-orfG.....T CA.....T..T.....C..T.	100
Consensus	AGTTCTTGAA GGRACTGAAW SMAAAATCAT YGCWAACCCA GAAGGMAAYC	100
spn-orf	.C.....A..T.....C.....C..A.....C.....T...	150
sga-orf	.T.....T..A.....A.....T..T.....T.....G...	150
sgb-orf	.T.....T..A.....A.....T..T.....T.....G...	150
Consensus	GYACAACTCC WTCWGTTAGTM TCATTCAAAA AYGGWGAAAT YATCGTKGGT	150
spn-orfT.....A..T.....C.....T..A..T.....	200
sga-orfC.....A..G.....C.....A..A..A.....	200
sgb-orfT.....G..A.....T.....T..T..T.....	200
Consensus	GATGCTGCAA AACGYCAAGC RGTDACAAAY CCAGAWACWG TWATCTCTAT	200
spn-orf	C.....T..G.....A.....	250
sga-orf	T.....T..A.....T.....	250
sgb-orf	C.....A..G.....T.....	250
Consensus	YAAATCWAAR ATGGGAACTT CTGAAAAAGT TTCTGCAAAT GGWAAAGAAT	250
spn-orf	.C.....A..C.....T.....C.....T..G.....C..C	300
sga-orf	.T.....T..T.....T.....A.....T.....C..T.....T..T	300
sgb-orf	.T.....T..T.....T.....A.....T.....C..T.....T..T	300
Consensus	AYACTCCWCA AGAAATYTCA GCWATGATYC TTCAATACYT KAAAGGYTAY	300
spn-orfC.....T..G.....ACC.....T.....C..A.....	350
sga-orfT.....A..A.....GAA.....A.....T..T.....	350
sgb-orfT.....A..A.....GAA.....A.....T..T.....	350
Consensus	GCTGAAGACT AYCTTGGWGA RAAAGTARMM AAAGCWGTTA TYACWGTTCC	350
spn-orf	G.....C.....C..T.....A.....A.....	400
sga-orf	A.....T.....T..A.....A.....T.....	400
sgb-orf	A.....C.....T..A.....G.....T.....	400
Consensus	RGCTTAYTTC AACGAYGCWC AACGTCARGC AACWAAAGAC GCTGGTAAAA	400
spn-orfT.....T.....C.....T.....A..T	450
sga-orfA.....C.....T.....A.....T..A	450
sgb-orfA.....C.....C.....C.....A.....C..A	450
Consensus	TTGCWGGTCT TGAAGTAGAA CGTATYGTTA AYGAACCAAC WGCAGCHGCW	450
spn-orfT.....A..A.....G..A..	500
sga-orfA.....G..T.....A..T..	500
sgb-orfA.....G..T.....A..T..	500
Consensus	CTTGCTTATG GTWTGGACAA GACTGACAAR GAWGAAAAAA TCTTRGTWTT	500

FIG. 24A

18/22

spn-orfC.....C..	T.....	..G.....C.	550
sga-orfT.....A..	A.....	..A.....T.	550
sgb-orfT.....A..	A.....	..A.....T.	550
Consensus	TGACCTTGGT	GGTGGTACAT	TYGACGTMTG	WATCCTTGAA	TTRGGTGAYG	550
spn-orfAT.GT..	..T.....G.	.C.....	600
sga-orfTC.TG..	..A.....T.	.T.....	600
sgb-orfTC.TG..	..A.....T.	.T.....	600
Consensus	GTGTCTTCGA	CGTWYTKKCA	ACWGCAGGKG	AYAACAAACT	TGGTGGTGAC	600
spn-orfA.....C..	...CCA...G	..A.CA....	.C.....	650
sga-orfA.....T..	...TTT...A	..G.CT....	.T.....	650
sgb-orfG.....T..	...TTT...G	..A.AA....	.C.....	650
Consensus	GACTTTGACC	ARAAAATYAT	TGAYYWCTTR	GTRGMWGAAT	TYAAGAAAGA	650
spn-orf	...C.....C	..CT.G..TA	CT..C..G..	...AA.G...	..T.....	700
sga-orf	...T.....T	..CT.A..AC	AA..T..G..	...AC.T...	..C.....	700
sgb-orf	...T.....T	..TC.T..TC	AA..C..A..	...TC.T...	..C.....	700
Consensus	AAAYGGTATY	GAYYTDTCWM	MWGAYAARAT	GGCWMTKCAA	CGYTTGAAAG	700
spn-orfG.....G..GC....	.T.....A..	TTC...A...	750
sga-orfT.....T..AT....	.A.....G..	ACA...A...	750
sgb-orfT.....T..AC....	.A.....A..	TCA...T...	750
Consensus	ATGCKGCTGA	AAAAGCKAAR	AAAGAYCTTT	CWGGTGTRAC	WYMAACWCAA	750
spn-orf	..CAGC..G.	.A..T.....	...A...GAGA....G..	800
sga-orf	..TTCA..A.	.G..C.....	...T...TCTT....A..	800
sgb-orf	..TTCA..A.	.G..C.....	...T...TCTT....G..	800
Consensus	ATYWSMTTRC	CRTTYATCAC	TGCWGGTKMK	GCTGGWCCTC	TTCACTTRGA	800
spn-orf	A....CT...	A.T.....G.T...	.T.G.....	850
sga-orf	G....GC...	T.T.....T.C...	.C.C.....	850
sgb-orf	G....GC...	T.A.....T.C...	.C.C.....	850
Consensus	RATGASYTTA	WCWCGTGCKA	AATTTGAYGA	TYTSACTCGT	GACCTTGTTG	850
spn-orfA..	.GT.....C.T...AGC	900
sga-orfG..	.AC.....T.A...TCA	900
sgb-orfG..	.AC.....T.C...TCA	900
Consensus	AACGTACRAA	ARYTCCAGTT	CGTCAAGCYC	TTTCAGATGC	AGGHTTGWSM	900
spn-orfC..C.....T...T....	.T.....	950
sga-orfT..T.....T...A....	.T.....	950
sgb-orfT..T.....C...A....	.A.....	950
Consensus	TTGTCAGAAA	TYGAYGAAGT	TATCCTYGTT	GGTGGWTCAA	CWCGTATCCC	950
spn-orf	T..C.....TT.C....	1000
sga-orf	A..A.....CA.T....	1000
sgb-orf	A..A.....TA.T....	1000
Consensus	WGCMTTGTY	GAAGCTGTWA	AAGCTGAAAC	TGGTAAAGAA	CCAAAYAAAT	1000

Fig. 24B

SUBSTITUTE SHEET (RULE 26)

19/22

spn-orf	.A..A.....AT....G.....T	1050
sga-orf	.T..A.....GT....T.....G	1050
sgb-orf	.T..T.....GC....T.....T	1050
Consensus	CWGTWAACCC	TGATGAAGTR	GTTGCIATGG	GTGCKGCTAT	CCAAGGTGGK	1050
spn-orf	..G..T....	.T.....C..	G..T.....CT.G.....	1100
sga-orf	..T..C....	.G.....G..	A..C.....CC.A.....	1100
sgb-orf	..T..C....	.G.....G..	A..C.....AC.A.....	1100
Consensus	GTKATYACTG	GKGATGTSAA	RGAYGTTGTM	CTTCTTGAYG	TAACRCCATT	1100
spn-orfC.....A..	A..T..A...	C.T.....T.	1150
sga-orfT.....T..	C..C..T...	T.G.....C.	1150
sgb-orfT.....T..	C..C..T...	T.G.....C.	1150
Consensus	GTCACCTGGT	ATYGAAACAA	TGGGTGGWGT	MTTYACWAAA	YTKATCGAYC	1150
spn-orfC..T..	A.....	1200
sga-orfT..A..	T.....	1200
sgb-orfC..A..	T.....	1200
Consensus	GCAAYACWAC	WATCCCAACA	TCTAAATCAC	AAGTCTTCTC	AACAGCAGCA	1200
spn-orfC...	1250
sga-orfT...	1250
sgb-orfT...	1250
Consensus	GACAACCAAC	CAGCCGTTGA	TATCCAYGTT	CTTCAAGGTG	AACGCCCAAT	1250
spn-orfG..T..	.T..A.....	1300
sga-orfG..T..	.T..T.....	1300
sgb-orfA..A..	.C..T.....	1300
Consensus	GGCAGCAGAT	AACAARACWC	TYGGWCGCTT	CCAATTGACT	GATATCCCAG	1300
spn-orfT..TC....	C.....C..G	1350
sga-orfC..AT....	T.....T..A	1350
sgb-orfC..AT....	T.....T..A	1350
Consensus	CTGCACCTCG	TGGAATYCCW	CAAATYGAAG	TAACATTTGA	YATCGAYAAR	1350
spn-orf	..C.....C.	.G.....T..	G..C.....C	..T..A..T.A.....	1400
sga-orf	..C.....T.	.T.....A..	A..T.....C	..T..T..G.G.....	1400
sgb-orf	..T.....T.	.A.....T..	A..T.....T	..C..T..T.A.....	1400
Consensus	AAYGGTATYG	TDTCTGTWAA	RGCIYAAAGAY	CTYGGWACKC	AAAARGAACA	1400
spn-orf	.ACT..T..C	...C....G.	.CTCA..TT.	GA....C...C..CC	1450
sga-orf	.CAC..C..T	...A....A.	.CGAC..AC.	TT....A...T..TC	1450
sgb-orf	.CAC..T..T	...C....T.	.TTCA..AT.	AA....T...T..TA	1450
Consensus	AMMYATYGTY	ATCMAATCDA	AYKMMGGWYT	DWCTGAHGAA	GAAATYGAYM	1450
spn-orf	GC.....	...T..A...	..A..C..T.	..T.C...AA	G.....	1500
sga-orf	GC.....	...C..T...	..T..T..C.	..G.C...GC	G.....	1500
sgb-orf	AA.....	...T..T...	..A..T..T.	..G.A...GC	A.....	1500
Consensus	RMATGATGAA	AGAYGCWGAA	GCWAAAYGCIY	AAKCMGATRM	RAACGTAAA	1500

FIS. 24C

20/22

sga-orfT. .C...AAA.. C.....T...	T....T..T..	1550		
sgb-orfT. .T...AAA.. T.....T...	C....A..A..	1550		
Consensus	GAAGAAGTWG	AYCTTMRWAA	YGAAGTKGAC	CAAGCHATCT	TTGCDACWGA	1550
spn-orf	...G..A..C ..G.....T.....	C..C..CG..T..C.	1600	
sga-orf	...A..A..C ..A.....T.....	C..T..CA..C..T.	1600	
sgb-orf	...A..T..T ..A.....C.....	T..T..TA..C..T.	1600	
Consensus	AAARACWATY	AARGAAACTG	AAGGYAAAGG	YTTYGAYRCA	GAACGYGAYG	1600
spn-orf	.T..C...G. T..C.....T	..CC.T..GA	AA..T.....	.GACAAC...	1650	
sga-orf	.A..G...T. A..T.....C	..GT.A..AG	CT..G.....	.TCTGGC...	1650	
sgb-orf	.A..G...T. A..A.....T	..GT.G..AA	AA..T.....	.TCAGGT...	1650	
Consensus	CWGCSCAAKC	WGCHCTTGAY	GASYTDAARR	MWGCKCAAGA	AKMHRRYAAC	1650
spn-orf	T.G.....A..A	T.G..C....T....G	1700
sga-orf	C.T.....T..A	T.A..T....G....C	1700
sgb-orf	C.T.....T..T	C.T..C....A....C	1700
Consensus	YTKGACGACA	TGAAAGCWAA	ACTTGAAGCW	YTDAAYGAAA	AAGCDCAAGS	1700
spn-orf	AC.T..T... ..C.C....	.A.....C..	A.....G...TCAAG	1750	
sga-orf	TT.G..T... ..A.G....	.G.....T..	A.....T...AGCAC	1750	
sgb-orf	TC.T..A... ..C.T....	.A.....G..	T.....A...AGCTC	1750	
Consensus	WYTKGCWGTT	AAAMBTBACG	ARCAAGCBGC	WGCAGCDCAA	CAAGCWSMWS	1750
spn-orfA..A.. ...C.....	G.AA-----	..GG..ACGC	A..C.....C	1794	
sga-orfT..A.. ...T.....	G.TAA.....	..GC..ATAA	---T....T	1797	
sgb-orfG..T.. ...T.....	T.AGC.....	..TC..GCAA	G..T....T	1800	
Consensus	AAGGDGCWGA	AGGYGCACAA	KCWRMTGATT	CAKSAARYRM	RGGYGATGAY	1800
spn-orf	..C.....C. .A..G..T..	G..A..G...			1824	
sga-orf	..T.....T. .C..A..T..	A..A..G...			1827	
sgb-orf	..T.....T. .C..A..C..	T..G..A...			1830	
Consensus	GTYGTAGAYG	GMGARTTYAC	DGARAARTAA		1830	

FIG. 24D

21/22

spn-prot	50	
sga-prot	50	
sgb-prot	50	
Consensus	MSKIIGIDLG	TTNSAVAVLE	GTESKIIANP	EGNRRTTPSVV	SFKNGEIIVG	50
spn-prot	100	
sga-prot	E.....	100	
sgb-prot	100	
Consensus	DAAKRQAVTN	PDTVISIKSK	MGTSEKVSAN	GKEYTPQEIS	AMILQYLKGY	100
spn-protT.....	150	
sga-prot	150	
sgb-prot	150	
Consensus	AEDYLGEKVE	KAVITVPAYF	NDAQRQATKD	AGKIAGLEVE	RIVNEPTAAA	150
spn-protL.....E.....S.....	200	
sga-prot	200	
sgb-prot	200	
Consensus	LAYGMDKTDK	DEKILVFDLG	GGTFDVSILE	LGDGVFDVLA	TAGDNKLGGD	200
spn-protH.....T....M.....S..	250	
sga-prot	250	
sgb-prot	E.....	250	
Consensus	DFDQKIIDFL	VAEFKKENGI	DLSQDKMALQ	RLKDAAEKAK	KOLSGVTQTQ	250
spn-protE.....T..T.....V..	300	
sga-prot	300	
sgb-prot	300	
Consensus	ISLPFITAGS	AGPLHLEMSL	SRAKFDDLTR	DLVERTKTPV	RQALSDAGLS	300
spn-prot	350	
sga-prot	350	
sgb-prot	350	
Consensus	LSEIDEVILV	GGSTRIPAVV	EAVKAETGKE	PNKSVNPDEV	VAMGAAIQGG	350
spn-prot	400	
sga-prot	400	
sgb-prot	400	
Consensus	VITGDVKDVV	LLDVTPLSLG	IETMGGVFTK	LIDRNTTIPT	SKSQVFSTAA	400
spn-prot	450	
sga-prot	450	
sgb-prot	450	
Consensus	DNQPAVDIHV	LOGERPMAAD	NKTLGRFQLT	DIPAAPRGIP	QIEVTFDIDK	450
spn-protT.....S.K...	500	
sga-prot	K...D..SE.....	500	
sgb-protK.....	500	
Consensus	NGIVSVKAKD	LGTQKEQHIV	IQSNSGLTDE	EIDRMMKDAE	ANAEADAKRK	500
spn-protR.....A.....A...D.....DN.	550	
sga-protA.....	550	
sgb-prot	550	
Consensus	EEVDLKNEVD	QAI FATEKTI	KETEGKGFDT	ERDAAQSALD	ELKKAQESGN	550
spn-protG.....	..QE.....	TGN.G-...	598	
sga-protM.....N...NN...	599	
sgb-protSA...SSKG..	600	
Consensus	LDDMKAKLEA	LNEKAQALAV	KLYEQAAAAQ	QAAQGAEGAQ	A.DSA...-DD	600
spn-prot	607	
sga-prot	608	
sgb-prot	609	
Consensus	VVDGEFTEK	609	

FIS 25

22/22

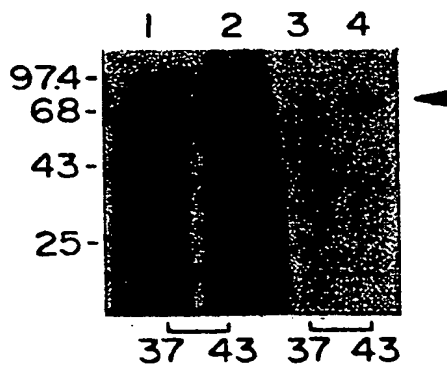


Fig. 26

INTERNATIONAL SEARCH REPORT

International Application No
PCT/CA 96/00322

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C07K14/315 C12N1/21 C07K16/12 A61K39/09 A61K39/395 G01N33/569 C12Q1/68		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C07K A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GENE, vol. 142, no. 1, 3 May 1994, AMSTERDAM NL, pages 91-96, XP002013303 M.J.S.BARRIL ET AL.: "Cloning and sequencing of the Lactococcus lactis subsp. lactis dnaK gene using a PCR-based approach" see figure 3	1,2, 6-21,23, 28,41
A	--- WO,A,93 17712 (BIOCINE SCLAVO SPA) 16 September 1993 see page 50	1-101
A	--- WO,A,92 14488 (UAB RESEARCH FOUNDATION) 3 September 1992 see the whole document -----	1-101
<input type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search <div style="text-align: center; font-size: 1.2em;">13 September 1996</div>		Date of mailing of the international search report <div style="text-align: center; font-size: 1.2em;">27.09.96</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016		Authorized officer <div style="text-align: center; font-size: 1.2em;">Cupido, M</div>

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CA96/00322

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 67,73,81,82,90,97-99
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 67,73,81,82,90 and 97-99 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the composition.
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/CA 96/00322

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9317712	16-09-93	CA-A- 2131551	16-09-93
		EP-A- 0632727	11-01-95
		JP-T- 7504423	18-05-95

WO-A-9214488	03-09-92	AU-B- 667668	04-04-96
		AU-A- 1445392	15-09-92
		AU-A- 4798096	27-06-96
		EP-A- 0571538	01-12-93
		JP-T- 6504446	26-05-94
		US-A- 5476929	19-12-95
